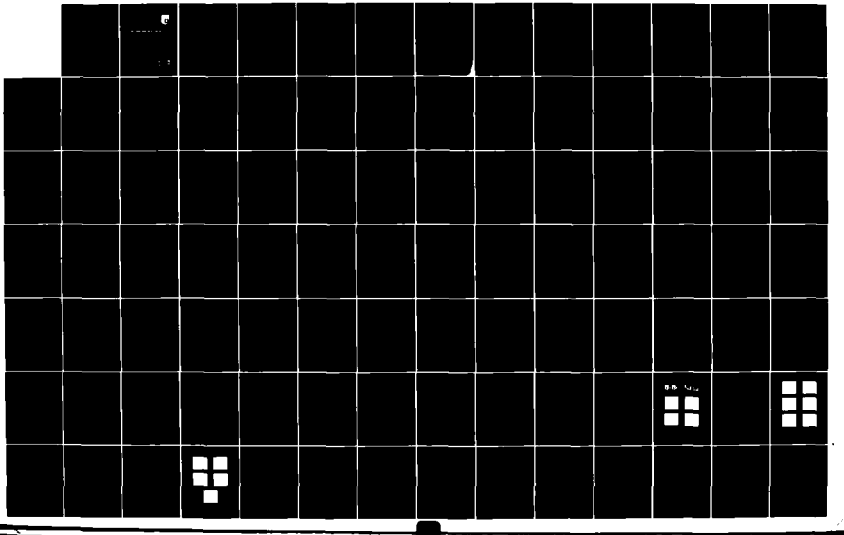
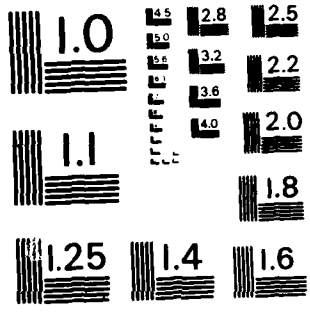
AD-A126 327 DIDA - DYNAMIC IMAGE DISPARITY ANALYSIS(U) MINNESOTA UNIV MINNEAPOLIS DEPT OF COMPUTER SCIENCE W B THOMPSON ET AL. 31 DEC 82 AFWAL-TR-83-1035 1/2

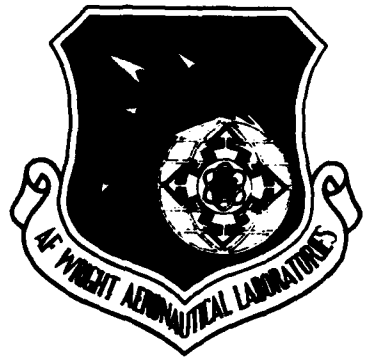
UNCLASSIFIED F33615-81-K-1541 F/G 9/2 NL





(12)

AFWAL-TR-83-1035



DIDA-DYNAMIC IMAGE DISPARITY ANALYSIS

WILLIAM B. THOMPSON
JOSEPH K. KEARNEY

COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

DECEMBER 31, 1982

Final Report for period July 1981 - December 1982

Approved for public release; distribution unlimited

AVIONICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Office of Public Affairs (ASD/PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

Louis A. Tamburino

LOUIS A. TAMBURINO
Project Engineer

Donald L. Moon

DONALD L. MOON, Chief, Information
Processing Technology Branch
Avionics Laboratory

FOR THE COMMANDER

Raymond D. Bellem

RAYMOND D. BELLEM, LT COL, USAF
Acting Deputy Chief
System Avionics Division
Avionics Laboratory

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AAAT-1 *, W-PAFB, OH 45433 to help us maintain a current mailing list".*

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFWAL-TR-83-1035	2. GOVT ACCESSION NO. AD-A126 327	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DIDA - DYNAMIC IMAGE DISPARITY ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Final Report July 1981-December 1982 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William B. Thompson Joseph K. Kearney		8. CONTRACT OR GRANT NUMBER(s) F33615-81-K-1541
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Minnesota 136 Lind Hall, Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ILIR 8109
11. CONTROLLING OFFICE NAME AND ADDRESS Avionics Laboratory (AFWAL/AAAT) AF Wright Aeronautical Laboratories, AFSC Wright-Patterson AFB, Ohio 45433		12. REPORT DATE 12/31/82 13. NUMBER OF PAGES 102
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image Understanding, Dynamic Image Analysis, Disparity Analysis, Optical Flow, Real-Time Processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Disparity is a point by point characterization of the translational changes in an image sequence due to motion of the sensor or objects under view. Accurate estimates of disparity are required in order to understand scene dynamics and to determine a wide variety of spatial relationships. This report describes the results of the first phase of the Dynamic Image Disparity Analysis (DIDA) project. The DIDA project was initiated in order to develop methods for estimating motion induced disparity in real-time.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The report outlines the types of approaches possible for estimating disparity. An extensive discussion on evaluating different approaches follows. Precise evaluation criteria are crucial if meaningful performance standards for DIDA are to be developed. The report argues that no single measure of accuracy is meaningful. Instead, performance should be characterized by a collection of measures. The evaluation of alternate disparity estimation techniques must consider this collection of measures together with a precise task analysis. In particular, different applications have different and often contradictory accuracy requirements.

Analytical and empirical analyses of a variety of potentially useful algorithms is described. Intrinsic error limitations of gradient based approaches are derived and this analysis is used to suggest improvements in the approach. Limited experimental comparisons of several different methods are presented. Architectural implications of certain promising methods are outlined. Finally, recommendations are made for continuing the DIDA program.

ACKNOWLEDGEMENT

We wish to thank the many people who assisted in the work described in this report. Daniel Boley helped us greatly in our understanding of the numerical analysis involved in the gradient techniques. Joel Neisen made substantial contributions to the empirical analysis and to the image processing and graphics system support. Stuart Levy wrote device drivers for us and brought up our image display particularly quickly. Kathleen Mutch and Richard Madarasz, while working on other projects for the University of Minnesota Computer Vision Group, provided helpful commentary on this project and supplied a variety of useful software. Robert Sedlmeyer and James Cohoon provided valuable criticism and important technical assitance. Finally, we would like to acknowledge the support and encouragement of Louis Tamburino who initiated the project and assisted us in both technical and administrative matters.

TABLE OF CONTENTS

1. INTRODUCTION

This report examines the feasibility of developing a device for the real-time estimation of motion induced disparity in image sequences. The report describes the nature of the disparity estimation problem and suggests criteria by which methods for estimating disparity can be evaluated. It includes a theoretical analysis of one class of estimation methods and shows how such an analysis can lead to improved performance. The results obtained from a variety of estimation algorithms are demonstrated on a limited sample of dynamic imagery. Finally, suggestions are provided for continuing activities in this program.

1.1. Motion Induced Disparity

Positional changes between an image sensor and objects in the environment can be described by using the concept of optical flow. The optical flow field specifies the instantaneous velocity on *the image plane* for every visible point on object surfaces. Non-zero values can occur due to object and/or sensor motion. Optical flow patterns can be used to estimate the direction of observer motion, the orientation of surfaces, the location of occlusion boundaries, and the relative distance to objects.

Since input is normally sampled at discrete moments in time, the image of a surface feature can translate significantly between frames. Borrowing from the terminology of stereo vision, we refer to this inter-frame translation as *disparity*. For purposes of interpretation, disparity values must be known over a reasonably dense sampling of image points. Thus, the disparity determination process must contend with the simultaneous tracking of a large number of points. Tracking only the most distinctive image features will not provide a sufficiently dense sampling of points. Consequently, the procedure which selects points to track must target imperfect feature points which may not be easily found in the subsequent frame or for which many ambiguous matches are possible.

The intrinsic difficulty of disparity determination is compounded by the need for real-time computation in many applications. Processing throughput of upwards of 2 million pixels per second may be required. Real-time implementation will be possible only if algorithms and processor architectures are carefully matched.

1.2. Background

While the determination of disparity in most realistic environments is extremely difficult, a number of reasonably successful systems have been demonstrated over the past several years. Three classes of approaches have been developed: temporal-spatial gradient analysis, matching, and differencing techniques. Each approach has both advantages and disadvantages with respect to effectiveness, generality, and efficiency.

1.2.1. Temporal-spatial Gradient Analysis

Temporal-spatial gradient analysis uses the change in intensity at an image point over both time and space to estimate the rate of translation of the underlying surface. It allows a point-by-point determination of disparity based on purely local criteria without the need to examine long sequences [1], [2], [3]. The process can be illustrated with a one-dimensional example. In figure 1.1, a surface characterized by an intensity wedge in the image is moving to the right. By measuring the slope of the wedge and the change in intensity at x_0, it is possible to determine the amount of translation. The technique can be extended to two-dimensional translation using several different techniques.

Temporal-spatial gradient analysis has been shown to be effective over a fairly broad range of imagery. It is reasonably efficient and hardware implementations for restricted forms have already been developed. Subpixel accuracy in determining the magnitude of disparity vectors may be possible. Problems include potential difficulties

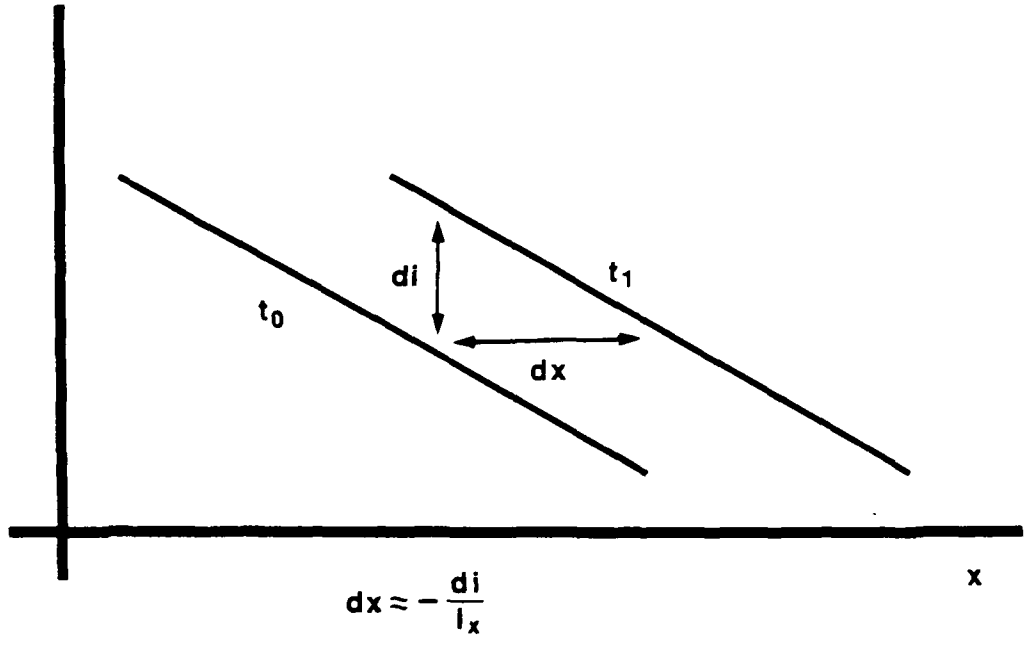


Figure 1.1 The intensity wedge represented by the diagonal line at t_0 moves dx along the x axis to t_1, causing a change in intensity of di.

in regions which large variations in disparity exist, errors generated by brightness variations due to causes other than motion, difficulty in accurately approximating the temporal and spatial gradients, and sensitivity to small measurement errors in regions of the image where the spatial gradients are nearly constant.

1.2.2. Matching

The most direct approach to the correspondence problem is matching. A set of structures is identified in one image frame and then an organized search for the corresponding structure is performed in subsequent frames. Usually, some optimization criteria is required in order find the "best" match for each structure. This may depend on properties of the structures, the relationships between structures, or both.

Matching can be done at the level of small image segments [4], [5], derived feature points [6], [7], [8], [9], [10], or image regions likely to correspond to full object surfaces [11]. The criteria functions can range from simple cross-correlation [4], [5], to complex and sophisticated graph-matching procedures [12].

The most straightforward matching techniques operate directly on intensity values. Individual, small regions are isolated in one image frame and used as templates. These regions may be located on a regular grid, but usually some attempt is made to choose only those regions that have a high information content, and hence a high probability of being accurately matched. For example, regions that have sharp autocorrelation peak can be chosen [13]. Corresponding regions are then searched for in a second frame using some form of cross correlation [13], [14], [15]. Such methods are computationally expensive. While some efficiencies are possible [16], much effort is still required to avoid false matches and to perform the matching at a dense enough sampling of points to be useful. Another problem with direct image matching is that the correct match for a template taken from one image may lie in a relatively distorted region in another image (because of surface relief, rotation, etc.). If this effect is to be accounted for, more degrees of freedom are introduced into the matching process and the computational efficiency suffers. Furthermore, the effect is more pronounced for larger templates.

As an alternative, it is possible to locate features points in an image and to match these feature points rather than the raw imagery [9], [10], [6], [7], [8]. Usable feature points range from local maximum of variability to image regions resulting from extensive segmentation operations. These *symbolic matching* techniques have several advantages. The amount of data to be processed can be significantly reduced. The number of possible matches which must be considered is often much less than that required for correlation based approaches, allowing both computational and representational efficiencies. Finally, carefully chosen feature points may minimize the effects

of luminance and geometric changes that can cause major difficulties for template matching systems. The major difficulty with symbolic matching is reliably determining structures to be matched.

1.2.3. Differencing

Differencing techniques start with a point by point determination of significant changes in image intensity. Often, this can be efficiently done by subtracting two image frames and thresholding the result. Clusters of points with above threshold differences correspond to portions of moving surfaces. The interiors of homogeneous image regions will not generate a difference, however, even if the corresponding surface is moving. Thus, longer sequences must be observed or more sophisticated, non-local analysis applied in order to determine surface boundaries. Once this is done, the rate of translation can be estimated by matching surfaces in different frames or by direct analysis of a sequence of difference pictures [17], [18], [19].

Differencing is a particularly efficient technique for dealing with some image sequences. The differencing operation itself is easily implemented. For a fixed sensor and an environment in which only a small portion of the scene is moving, no change will be evident over most of the image. By concentrating only on differences, significant data reduction is possible. Important limitations include difficulties with situations in which most or all of an image is changing over time (such as with observer motion), problems with occlusion boundaries between two moving objects, and imprecision in the disparity estimates. This in turn limits the generality of the differencing approach and it will not be discussed further in this report.

1.2.4. Interpretation

Disparity fields provide important information about the relative spatial position and velocity of a sensor and visible objects and surfaces. If the sensor is following a

know trajectory through an otherwise static environment, precise three-dimensional shape information about that environment can be obtained. If moving objects are present, some shape properties can still be determined even though the actual distance from sensor to object is no longer easily computable. With minimal knowledge about object type and motion, disparity analysis still allows target cueing and object boundary identification.

If a sensor moves with known velocity along a known path, simple trigonometric relationships can be used to determine distance to a surface directly from disparity (eg. [20], [21]). This "motion stereo" technique has been extensively studied, though as yet no procedures exist for rapidly computing disparity values with sufficient precision to allow accurate depth estimation over a broad range of scene types.

Many of the most important applications of disparity analysis arise in situations in which sensor trajectory and/or velocity is not known with precision. For example, if a sensor is not rotating, the relative orientation between the sensor's optical axis and direction of travel can be found by locating the focus of expansion of the disparity field. In the same situation, the orientation of visible surfaces relative to the sensor can be found, even though nothing is known about sensor velocity [22]!

In even less constrained situations, it is still possible to use disparity values to locate occlusion boundaries and thus find object boundaries and depth discontinuities. Gradual changes over space in disparity correspond to continuous surfaces while abrupt changes correspond to depth discontinuities. Additional analysis allows determination of which side of the boundary corresponds to the occluding surface.

1.3. Need For Continued Study

In order to design and construct a disparity estimation system, three aspects of dynamic image analysis must be studied: effectiveness, generality, and efficiency. In addition, efforts must be made to understand the interrelationships between these pro-

perties.

Effectiveness is a measure of the accuracy and utility of the analysis system. Methods are required which estimate optical flow with greater precision and at a denser sampling of points on the image plane than is currently possible. Limits on accuracy of interpretation processes must be studied. Mechanisms for systematically evaluating effectiveness should be created.

The *generality* of existing methods is seldom discussed explicitly. In fact, all methods presume constraints upon scene type, scene dynamics, camera model, and image properties. As an example, some techniques work only for a fixed sensor and moving objects while others work only for a moving sensor in an otherwise stationary environment. A better understanding of the need for these limiting constraints is required.

Efficiency is obviously important if dynamic image analysis is to be used in any real-time applications. Two forms of efficiency must be considered: throughput (the input data rate that can be accommodated) and latency (the time between activity in the scene and the production of a description of that activity. Efficiency is not just an issue of clever hardware design. Algorithms must be structured in a "conceptually efficient" manner so that they easily map onto appropriate computer architectures.

2. EVALUATION

It would be most desirable to find an estimation technique which would be able to calculate disparity at every point in the image with arbitrary precision. Unfortunately, the ambiguity of the problem, noise in the data, and the restrictions of time and space make it unreasonable to expect anything approaching perfect performance. Accepting that errors are inevitable, the next problem is to find the estimation technique which produces the best results and to judge whether or not performance is satisfactory to accomplish a desired goal. The definitions of the best result and satisfactory performance will depend upon the objectives of the system. For example, some navigation tasks require that disparity be know very accurately but only at a small number of points. Other tasks, such as segmentation, require a dense sampling of disparity vectors which need be known with relatively little accuracy. Thus, the criteria which are used to evaluate disparity estimation techniques must depend upon the task to be accomplished.

The performance of a technique will be affected by the environment in which it is used. The nature of surface reflectance, the number and size of objects in the scene, and the characteristics of the motion can all affect the quality of the results. Some techniques are designed for constrained motions and will only work only in very special environments. Before an estimation technique can be evaluated, the environment in which the technique will operate must be specified.

The problem domain will determine both the task to be accomplished and the environment in which the task will be performed. The relationship between the problem domain and performance evaluation is diagrammed in figure 2.1. Throughout this report we will be examining the performance requirements demanded of disparity estimation techniques by application tasks and the performance dependencies of estimation techniques.

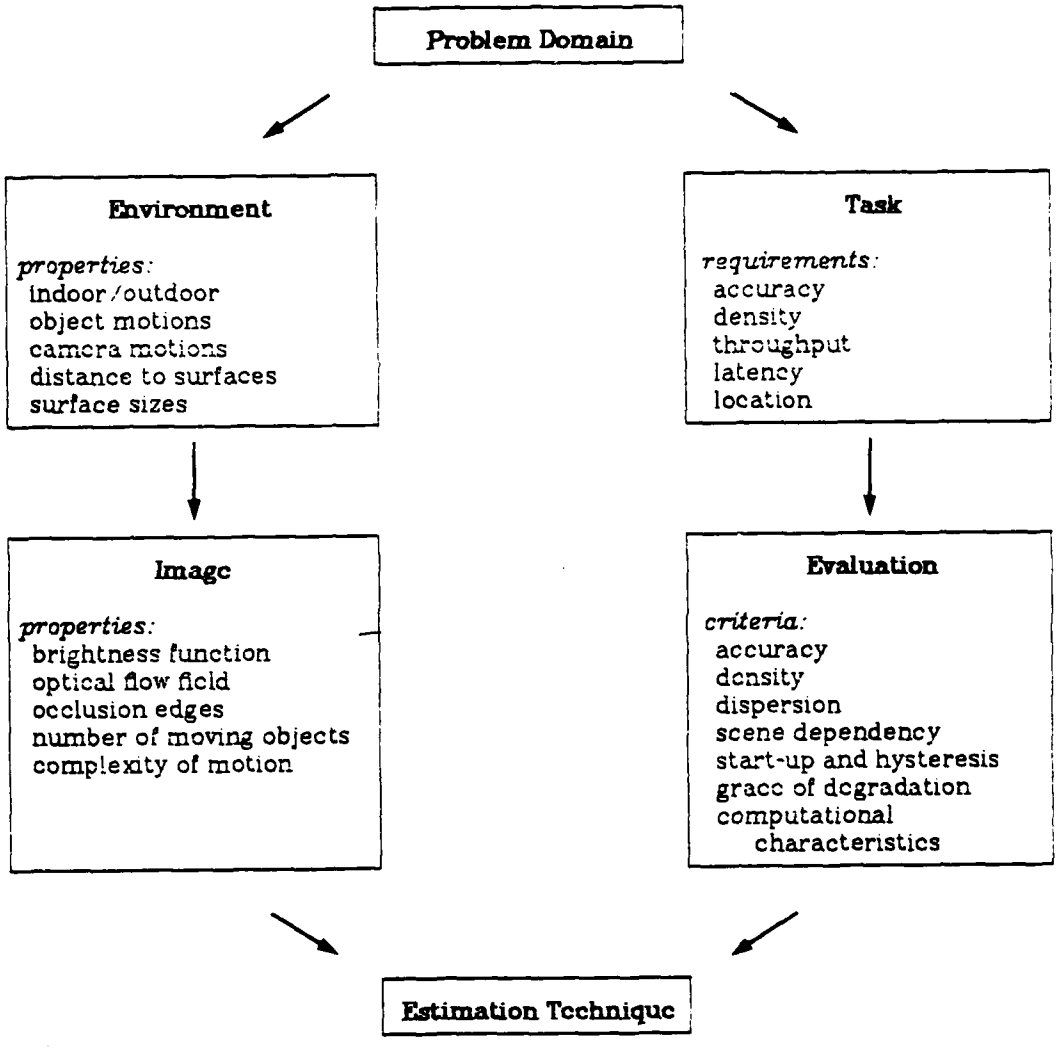


Figure 2.1 The relationship between the problem domain and evaluation of disparity estimation techniques.

In this section we will examine the nature of the information which is available from optical flow and identify the features of optical flow estimates which are required to retrieve this information. By way of example we will introduce three specific problems for which optical flow can provide important information and describe the charac-

teristics of the estimation technique which are important to each application.

Through our study of the intrinsic information content in flow fields we will infer the important properties of flow field estimates which determine the capability of interpretation methods to retrieve information about the scene. We will arrive at a set of characteristics which describe the important aspects of performance for estimation techniques. Each application which is based upon optical flow will depend upon a subset of these performance features.

2.1. Goal Dependency

Optical flow can provide useful information about the structure and dynamics of a three-dimensional scene. The optical flow field is determined by the positions and positional changes of objects in the scene relative to the camera. The extent to which this information is recoverable depends upon the constraints which can be placed upon environment. If an accurate camera model is available and sufficient information about the movement of objects and the sensor is available, depth relationships may be found by triangulation. Likewise, if there exists an accurate description of the spatial layout of a scene, the location and motion of the camera can be specified.

Important information about position and motion can also be obtained in less constrained situations. If an observer is moving through a static world both the direction of motion and the orientation of visible surfaces are computable without knowing the observer's speed. Even when moving objects are present, successive views of a scene can yield information about the location and nature of occlusion boundaries. Under relatively general viewing conditions moving targets can be located against a stationary background, objects which lie on a collision path with the observer can be detected, and the time to collision can be estimated.

Interpretation of optical flow to recover motion and shape information requires that optical flow be estimated from a sequence of discrete images. The knowledge

which can be gained from optical flow field estimates directly depends upon the performance characteristics of the estimator. In the next section we identify a set of features which characterize the important aspects of optical flow estimation. Specific methods are introduced to measure the performance of estimation techniques with respect to some of the features. For other features more qualitative judgements are suggested. Each feature is important for some interpretation task but may not be relevant to others. The feature set, by itself, provides a descriptive tool for characterizing the performance of disparity analysis techniques. When coupled with a problem specification in which the requirements of the estimator are clearly delineated, the feature measurements can be used to evaluate the appropriateness of estimation techniques.

Before introducing the feature set we will consider three applications areas for which optical flow estimation is important. This analysis will illustrate the importance of a variety of features and the differential requirements of different problem areas.

2.1.1. Navigation

Optical flow has proven to be useful for autonomous vehicle guidance in a variety of contexts [23,24,25]. If the environment is familiar, optical flow can be used to orient the vehicle with respect to a map of the environment. When an accurate map of the environment is not available, optical flow can be used to create a three-dimensional description of the area under view. This can be used to create a map, to dynamically determine goals (eg. possible targets), or to avoid obstacles.

The location and velocity of a moving observer can be computed from the disparity of a small number of *control points* in dynamic imagery. The position of the control points must known very accurately in the maps coordinate system. Determination of position and velocity with this method requires prior knowledge about the topography of the environment at selected points and the ability to identify map control points in

the image.

If the three-dimensional structure of the environment is highly varied, as in an urban environment or in rugged terrain, then three-dimensional descriptions of the scene can be matched with a three-dimensional model of the environment to determine the location and velocity of the sensor. Assuming that the vehicle is traveling through a static environment, depth relations in the scene can be computed from flow vectors. The accuracy requirements here are much less stringent than for the control point method presented above, however, a much more dense sampling of estimates is required. Distinctive features in the depth map, constructed from the image sequence, can be compared with a stored representation of the environment to place the vehicle within a model of the environment.

Even where little is known about the makeup of the environment, optical flow can provide important information for guidance and navigation. Consider the problem of maneuvering a land based vehicle through an obstacle course. All that is known is that the vehicle lies on a relatively flat surface on which also lie a number of stationary rigid objects. A rover on mars might face such a situation [24]. If the vehicle can monitor its motions by means of an inertial guidance system or if as in [24] camera motions are independently manipulated while the rover is stationary, then the distance to surfaces can be computed. This information can be used to move through the environment and to construct a model of the environment.

In situations where observer motions can not be determined, optical flow can still be used to make *qualitative* inferences about the structure of the environment [26,27]. The location surface boundaries, the relative positions of objects, the number of objects, the positions of obstacles with respect to the path of the vehicle, and the time to collision of with objects in the path of a moving camera are all potentially derivable from optical flow. Discontinuities in the flow field correspond to depth discontinuities. Occlusion boundaries can be found by examining the discontinuities in the flow field.

This requires that disparity be known near the occlusion boundaries in the image. Expansion and contraction of the flow field is related to the approach and withdrawl of surfaces. This information is critical to avoid collisions. To assure that no objects in the field of view escape the attention of the collision avoidance system, disparity must be calculated over the entire image.

The accuracy required of optical flow determination for use by qualitative techniques is not yet well understood. Some preliminary results indicate that it may be desirable to integrate the estimation of disparity with the interpretation of spatial relationships when qualitative information is sought [27]. In such a situation, it is more sensible to talk about the accuracy of the estimated spatial properties than the accuracy of the optical flow estimates.

2.1.2. Terrain Mapping

Aerial photographs can be used to automatically map the topology of the ground below. If the position and orientation of the camera at the time each image was taken are known, then the depth to each point in the image can be obtained by triangulation. Elevation maps are conventionally constructed by hand, matching points between temporally adjacent pairs of images from an aerial sequence. The matching process is time consuming, expensive, and error prone. Optical flow estimation techniques offer the potential to automate a large portion of this work.

The camera position and orientation can be derived from the disparity of ground control points. Markers which are highly distinctive (in the image) are identified on the ground. These markers may be a natural part of the scene, such as a radio tower, or may be placed there specifically for the camera calibration. The coordinates of the markers must be precisely measured on the ground. The camera position and orientation can be calculated by reversing the procedure used for elevation determination. Knowing the locations of a small number of control points in two images and the posi-

tions of the control points in some coordinate scheme allows the position and orientation of the camera to be obtained by triangulation. This process is essentially the same as was encountered in navigation when solving for observer position and motion.

While both camera modeling and elevation estimation depend upon disparity estimates, the two processes place very different requirements on the estimation techniques. Disparity need be known at only a small number of points in order to solve for the camera model and these points are usually chosen to be highly distinctive and hence, easy to match. Because the solution methods for calculating the camera model tend to be ill-conditioned, disparity must be know very precisely at the control points.

In contrast, detailed mapping requires that disparity be known over most of the image. This necessitates the estimation of disparity at a dense sampling of points. In regions where disparity can not be well estimated at each point, estimates must be well distributed so that there are not large regions of the image for which elevation is not known. While the accuracy of the depth measurements is dependent upon the accuracy of the disparity measurement on which it is based, the calculations are much less sensitive to errors than the camera modeling schemes.

2.1.3. Target Cueing

Most applications of optical flow analysis depend in large part on inferring depth relationships in the scene under view. Optical flow analysis may also be used, however, to locate areas of potential interest in an image without any direct concern for determining actual spatial relationships. In particular, optical flow analysis is useful for cueing on targets moving against a stationary background. This is possible even if the sensor is moving, resulting in a continually changing image of the background.

There are two possible approaches to dynamic target cueing using optical flow information. The optical flow field can be used to register subsequent frames to facilitate simple *change detection* techniques. The flow fields may also be used directly for

moving object detection.

A variety of techniques have been proposed for detecting differences in two or more image frames (eg. [28]). Most use the same basic approach. First, the frame are registered with one another. This is usually done by identifying a set of recognizable ***tie points*** in each image. The tie points are used to solve for an interpolation function that can be used to map all images into the same spatial coordinate system. Once this registration has been performed, some type of pointwise difference function is applied to pairs of frames in the sequence. The difference function can be as simple as the subtraction of the corresponding pixels, followed by a search for differences above some threshold. Often, preprocessing and/or more sophisticated tests for significant differences are used.

A primary advantage of this change detection approach is its computational simplicity which can be exploited to ease problems with real-time implementations. There are several difficulties with the approach. however. Tie-points must be known with great accuracy. This may not be possible if there are few highly distinguishable points in view or if the background is moving rapidly relative to the sensor. At best, only a sparse sampling of tie-points is usually available. For this reason and because of the desire to limit computational complexity, interpolation is usually performed using a low order polynomial function. Such interpolation techniques are likely to result in relatively large errors in areas of the image where disparities are changing rapidly due to changes in depth in the original scene. As a result, many false positive responses are likely to occur in these regions.

An alternate approach is to use optical flow analysis and a moving sensor to directly locate moving targets. If a sensor moves over a static background, the resulting disparity field will appear to expand radially from a point known as the ***focus of expansion*** (FOE). Given a flow field, it is relatively easy to solve for the location of this point in image coordinates. If there are moving objects in the field of view, but they

make up only a relatively small portion of the image, then it is still possible to solve for the FOE in a robust manner. Knowing the FOE places a directional constraint on the disparity of all points associated with the background. Any disparity not radiating out from the FOE must be associated with a moving object. The major difficulty is for objects with motions that coincidentally satisfy this constraint. In most situations, the constraint will only be satisfied momentarily. Motion of the sensor will quickly lead to detection of the target. It is also possible to look for rapidly changing disparity values along lines radiating from the FOE. If the change is too large to be accounted for by an expected change in depth, then it is due to a moving object.

2.2. Evaluation Criteria

In the last section several specific problems which make use of optical flow where introduced. Consideration of the requirements of different applications demonstrated how interpretation algorithms depend upon the performance of estimation techniques. The different requirements are summarized in table 2.1. The important characteristics of optical flow field estimates are surveyed in this section.

2.2.1. Accuracy

The accuracy of disparity estimates over an area of the image S can be determined from the average error in the estimates. The expected error is simply

$$E(\varepsilon_S) = \frac{\sum_S \| \hat{\mathbf{w}}_{ij} - \mathbf{w}_{ij} \|_2}{\sum_S 1} \tag{2.1}$$

where $\hat{\mathbf{w}}_{ij}$ is the estimate of disparity at the point (i,j) and $\mathbf{w}_{ij}$ is the true value of disparity:

$$\hat{\mathbf{w}}_{ij} = \begin{bmatrix} \hat{u}_{ij} \\ \hat{v}_{ij} \end{bmatrix} \quad \text{and} \quad \mathbf{w}_{ij} = \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} \tag{2.2}$$

Problem	Interpretation Task	Requirements		
		density	accuracy	location
navigation unknown environment	point matching	low	high	control points
	depth matching	high	moderate	distinctive regions
navigation known environment	collision avoidance	high	unknown	everywhere
	occlusion boundary detection	moderate	unknown	near occlusion edges
mapping	camera model	low	high	control points
	surface elevation	high	moderate	everywhere
cueing	change detection	moderate	moderate	target and background

Table 2.1 Characteristics of optical flow field estimates required by different application tasks.

A major difficulty in estimating the accuracy of a disparity field is to determine the ground truth against which the estimates are to be compared. The acquisition of ground truth data is discussed in detail in the next section

Frequently, an index of the accuracy of an estimate is generated with the disparity estimate. This index can be treated as the confidence that the estimator has in its estimate. It would be desirable to consider the additional information which confidence provides when evaluating the accuracy of a disparity field. One measure of accuracy which reflects the confidence in the estimates is the average of the estimation errors weighted by the confidence in the estimate:

$$E(\varepsilon_s) = \frac{\sum_S p_{ij} \| \hat{\mathbf{w}}_{ij} - \mathbf{w}_{ij} \|_2}{\sum_S p_{ij}} \tag{2.3}$$

where p_{ij} is the confidence in disparity estimated at the point (i,j).

It is also important to know how well a technique is able to judge its own performance. This requires that the association between the confidence estimate and the true error in the disparity estimate be specified. Confidence can be interpreted as a measure of the expected accuracy in the estimate. At low values of confidence the magnitude of the error is likely to be large. One estimate of the effectiveness with which confidence predicts accuracy is the correlation between the expected error $\bar{\varepsilon}$ and confidence. To calculate an expected error for values of the continuous confidence variable p, we divide confidence into n subranges, $p(1), p(2), \ldots, p(n)$. Let $\bar{\varepsilon}_{p(i)}$ be the expected error for estimates within the subrange of confidence $p(i)$. The correlation between the confidence and the expected error in the disparity estimate is given by

$$r = \frac{n \cdot \sum_{i=0}^{n} \bar{\varepsilon}_{p(i)} p(i) - \left[\sum_{i=0}^{n} \bar{\varepsilon}_{p(i)}\right] - \left[\sum_{i=0}^{n} p(i)\right]}{\sqrt{n \cdot \sum_{i=0}^{n} \bar{\varepsilon}_{p(i)}^2 - \left[\sum_{i=0}^{n} \bar{\varepsilon}_{p(i)}\right]^2} \sqrt{n \cdot \sum_{i=0}^{n} p(i)^2 - \left[\sum_{i=0}^{n} p(i)\right]^2}} \tag{2.4}$$

For some applications only one component of disparity is relevant. An observer translating through a static environment can determine the distance to surfaces knowing only the magnitude of disparity. In the same context, the focus of expansion of the flow field can be estimated from the orientation of disparity vectors. When attempting

to distinguish a moving object against a large background, small changes in vector orientation are much more significant than small changes in the magnitude of disparity. To judge the appropriateness of disparity analysis techniques for these applications, accuracy measurements should be resolved into orientation and magnitude components.

2.2.2. Density and Dispersion

The density of the disparity field and the dispersion of the estimates across the field are important characteristics of an estimator. The density of disparity estimates over an area S on the image can be found by

$$\rho = \frac{\sum_S B(i,j)}{\sum_S 1} \tag{2.5}$$

where,

$$B(i,j) = \begin{cases} 1 & \text{if disparity estimated at } (i,j) \\ 0 & \text{if disparity not estimated at } (i,j) \end{cases}$$

In this formulation ρ varies between 0 (no estimates are present in S) and 1 (disparity is estimated at every point in S).

Measuring density by (2.5) assumes that at each point a disparity estimate is either available or not. This approach does not take into account the certainty of the estimate. The density of the estimates can be treated as a measurement of the amount of information which the estimation technique has extracted about an area. Another way to capture the amount of information acquired over an area is to examine integral of confidence over the area. If the confidence estimates are closely associated with the likelihood that an estimate is in error then the knowledge which is contained in a field of estimates can be measured by

$$\rho = \frac{\sum_S p_{ij}}{\sum_S 1} \tag{2.6}$$

where, as before, μ_{ij} is the confidence in $\hat{\mathbf{w}}_{ij}$

In many circumstances the dispersion of disparity estimates is as important as the accuracy or density of estimates. If all of the knowledge which is acquired from the image is concentrated in a few small areas then little can be inferred about the three-dimensional structure of the image. Dispersion characterizes the way in which the estimated disparity vectors are distributed over the image. As a trivial example of the importance of dispersion, consider the value of a system only capable of estimating disparities in the upper right corner of the image.

For many applications, it is desirable that estimated values be distributed in an approximately uniform fashion over the field of view. This property can be described by a measurement of *unconditional dispersion*. The dispersion of disparity estimates is strongly tied to the dispersion of motion information. Thus, the dispersion of estimates is likely to be quite sensitive to the characteristics of the imagery. Textureless regions contain no information about motion. Consequently, it would be expected that estimates, for any disparity analysis technique, would be concentrated in more textured areas. The dispersion of the vectors will be highly dependent upon the amount of texture and dispersion of texture in the image. Statistical characterizations such as entropy can be used to quantify the unconditional dispersion of disparity estimates.

The effectiveness of some forms of analysis depends heavily on the *conditional dispersion* of estimated values. For example, if it is important to find depth discontinuities which correspond to object boundaries, then evaluation techniques must determine the dispersion of estimated disparity values in the vicinity of all such boundaries in the original scene. Conditional dispersion is more difficult to quantify. Its measurement depends on the availability of a complete task description along with a model of the scene under view.

2.2.3. Scene Dependency

Some problems require that optical flow be known only at specific locations in the image. For example, to segment the image into continuous surfaces it is necessary that disparity estimates be obtained near occlusion boundaries. Disparity need be known only at a some set of control points to calculate camera motion parameters. It is important to understand how disparity estimation techniques perform in semantically important regions of the image.

Frequently, the environment in which the disparity is to be estimated is very stereotyped. The size, shape, and reflectance characteristics of the objects which are to be observed may be known in advance. The camera and objects motions may be limited. The lighting conditions could be adaptive under the control of the observer. In order to take advantage of these constraints and to understand the limitations which these constraints place upon disparity estimation it is necessary to characterize how performance depends upon the viewing context. Among the important features of the environment are:

1. The number and size of the objects in the scene.
2. The reflectance characteristics of the objects and the background -- most importantly, the amount of texture.
3. Allowable camera motions.
4. Allowable object motions.

The list above should not be considered as comprehensive. As problem domains are better understood new aspects of the environment are likely to prove to be as important as those above.

2.2.4. Start-up and Hysteresis Characteristics

The population of visible object points is continually changing in a dynamic scene. Points appear and disappear at the border of the image as new objects enter and leave the field of view. Surfaces occlude and disocclude the objects behind them. The appearance and disappearance of object points in the scene can be a source of information about the scene. Examination of the regions of accretion and deletion which surround moving objects can be used to determine which surface is in front of the other. Unfortunately, the appearance and disappearance of object points can also confound techniques which estimate optical flow.

Abrupt changes in scene dynamics can also be a potential source of difficulty. Rapid accelerations and decelerations, as might be observed when two moving objects collide, and sudden changes in view can lead to significant breakdowns in disparity estimation.

The manner in which techniques respond to changes in the makeup of the image and changes in scene dynamics are important features of disparity estimation techniques. Both the magnitude of the degradation in performance and the length of time which it takes to recover must be considered. For techniques which depend upon the output of one stage to initialize the processing in the next time interval the start-up characteristics should be understood.

2.2.5. Grace of Degradation

The manner in which techniques fail is an important characteristic of their performance. If procedures are robust, small changes in the scene should not lead to a significant deterioration in performance. However, robustness is a difficult trait to quantify. One way to roughly gauge the grace with which a procedure fails is to slowly introduce noise into the scene and observe the performance degradation. Here, noise is taken as any facet of the dynamic scene which is know to degrade performance. The

rate at which performance declines is an indication of the robustness of the procedure.

When performance deteriorates it is often helpful to be aware of the degradation. Many disparity analysis techniques provide a measure of the quality of estimates. Such measures of confidence can be used by higher level processes to adapt to the loss in performance. Serious errors in interpretation can be avoided by delaying judgements until better estimates are available or by basing interpretation upon earlier estimates. Techniques for evaluating confidence estimation procedures were discussed in section 2.2.1. The importance of accurately measuring confidence is that the affect of breakdowns in performance can be minimized.

2.2.6. Computational Characteristics

Many applications of dynamic image analysis must be performed in real-time. If real-time performance is to be obtained the amount computation and the nature of computational processes which can be performed must be severely constrained.

A variety of architectures have been recently developed to implement image processing algorithms. (See, for example, [29].) The degree to which algorithms can be mapped into feasible architectures will determine their suitability for real-time applications.

Algorithms with a high degree of parallelism are well suited to real-time architectures. Image processing necessitates the processing of a large amount data. In order to efficiently process whole images, algorithms must be structured as a large number of local independent processes. Interactions among processes should be minimal and highly localized.

Computations which are necessarily serial are best structured as a sequence of independent computations. If each successive operation is not contingent upon the results of prior or future processing then the computations can be adapted to pipeline architectures.

The ability to decompose an estimation technique into parallel algorithms and pipeline processes is an important characteristic for real-time applications.

2.3. Determining Ground Truth

In order to determine the accuracy of disparity estimates the "true" optical flow field must be known. The accuracy of the ground truth data limits the quality of the evaluation -- estimates can be judged to be no more accurate than the standard against which they are compared. As might be expected, determination of ground truth optical flow is a very difficult problem, otherwise the approximation techniques described in this report would not be of interest. Disparity can be estimated in a variety of ways which are practical for special circumstances or which are prohibitively expensive or too time consuming for general use. These techniques can, however, be used to validate more general purpose approaches.

In real-world environments, optical flow can be computed if the geometry of the environment and the photographic conditions are well known. The location of an object point in the image can be predicted from the three-dimensional position of a visible object point, the location and orientation of the camera, and the optics of the camera system. For most environments, precise three-dimensional position is available for only a few points in the field of view. Consequently, disparity can only be determined for a small number of points.

For some indoor problem environments representative scenes can be selected and the camera and object positions can be measured. It is also possible to construct physical models of environments in which it is unreasonable to perform positional measurements directly. The simulated environment must be realistic, containing all of the potential difficulties which might be encountered in the problem environment. The ground terrain belt developed at Wright-Patterson Air Force Avionics Laboratory is a good example of a simulated environment.

Range sensing devices such as radar or sonar can be used to acquire a depth map of the scene, where they are available. Alternatively, depth can be estimated by means of artificial illumination, commonly called *structured light*. Grid projection and lasar light sources are are frequently used to calculate range by triangulation between the light source and a single camera. Structured light can also be used to simplify disparity estimation with two cameras. Illumination of single points in the scene trivializes the identification of correspondences in the frame pair. Similarly, line or grid projection greatly simplifies the correspondence problem.

Disparity can be estimated by visual inspection of a frame pair by a human observer. Topological maps are commonly obtained from aerial imagery in this manner [30]. Depth may be judged monocularly, by matching points between images on their two-dimensional appearance, or binocularly, by using a stereoscopic display. The process is too time consuming for most applications but offers potential for the collection of ground truth data.

Optical flow need not be known at every point in the image to evaluation disparity analysis techniques. An accurate estimation of performance can be obtained with only a small number of points by using statistical polling techniques. The population of points must by sufficiently large to contain a representative sampling of values across the range of important environmental and image properties. Alternatively, the sample of points could be selected to cover the range of important environmental and image properties where they are known.

The accuracy of the ground truth estimates can be improved, relative to the estimates of approximation techniques, by basing the ground truth upon an oversampled image. The performance of disparity analysis techniques on an undersampled data set can be compared to ground truth estimates based upon the more densely sampled image.

Ground truth determination is a significant unsolved problem. No current method is completely satisfactory. Collection of a standard data base of image sequences with accurate ground truth measurements would make an important contribution to the vision research community.

3. AN ANALYSIS OF THE GRADIENT-BASED APPROACH

The choice of a disparity analysis algorithm should depend upon the performance characteristics of the estimation technique and the performance requirements of the task to be performed. For a variety of problems, gradient-based methods offer significant advantages over matching techniques for estimating disparity. The most salient difference between matching and gradient-based approaches is the density of points on the image plane at which disparity can be estimated. Matching techniques are highly sensitive to ambiguity among the structures to be matched. Disparity can be accurately estimated for only highly distinguishable regions. This means that disparity can only be determined at a sparse sampling of points across the image. Furthermore, it is computationally impractical to estimate matches for a large number of points. The gradient-based approach allows disparity to be simply computed at a more dense sampling of points than can be obtained with matching methods.

Gradient-based techniques avoid the difficult task of finding distinguishable regions or points of interest. The gradient approach leads to algorithms which are characterized by simple computations localized to small regions of the image. These techniques can be applied over the entire image. As we shall see in the analysis that follows, the gradient technique is also sensitive to ambiguous areas – no technique can locally determine the motion of a homogeneous region. The loss of precision in ambiguous areas can be quantified. This allows poor estimates to be filtered from the flow field. The measurement of the accuracy of disparity estimates can be obtained as a by-product of the estimation process and requires little additional computation.

Gradient-based techniques offer the additional advantage that estimates can potentially be made with sub-pixel accuracy without resorting to complex interpolation functions. A third advantage to gradient-based techniques is that the computational structure is simple and may be adapted to special purpose architectures.

The gradient-based approach for estimating disparity has been widely studied. [2,1,3,31,32,33,34,35,36]. A number of algorithms have been proposed with variations and enhancements to improve performance. This section examines the causes of error and the error propagation characteristics of one class of gradient-based algorithms. By understanding how errors arise in disparity estimates we are able to define the inherent limitations of the technique, obtain estimates of the accuracy of computed values, enhance the performance of the technique, and demonstrate the informative value of some types of errors.

3.1. The Gradient Constraint Equation

The gradient constraint equation can be derived as a Taylor series expansion of the image brightness function. It is assumed that the observed brightness (intensity on the image plane) of any object point is constant. Let brightness at a point $\mathbf{p} = (x,y)$ on the image, observed at time t, be represented by $I(x,y,t)$. Consider a point which is displaced by the vector $(\delta x,\delta y)$ over the interval δt:

$$I(x,y,t) = I(x+\delta x,y+\delta y,t+\delta t) \tag{3.1}$$

Following [31] we expand the image brightness function in a Taylor's series around the point (x,y,t) to obtain

$$I(x+\delta x,y+\delta y,t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + \varepsilon \tag{3.2}$$

where the remainder, ε, consists of the higher order terms of the expansion. Assuming that δx and δy vary with δt we can express ε as $O(\delta t)$. Subtracting $I(x,y,t)$ from both sides of (3.2) and using the constant brightness assumption formalized in (3.1) we have

$$0 = \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + O(\delta t) \tag{3.3}$$

To find a expression which relates velocity on the image plane to the gradients of brightness we divide (3.3) by δt and obtain

$$0 = \frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} + O(t) \tag{3.4}$$

Taking the limit of (3.4) as $\delta t \to 0$, we arrive at the gradient constraint equation:

$$0 = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} \tag{3.5}$$

The gradient constraint equation relates velocity on the image plane to the spatial and temporal gradients of brightness. For convenience we will make the following notational substitutions:

$$u = \frac{dx}{dt}, \quad v = \frac{dy}{dt}$$

$$I_x = \frac{\partial I}{\partial x}, \quad I_y = \frac{\partial I}{\partial y}, \quad I_t = \frac{\partial I}{\partial t}$$

The gradient constraint equation can now be stated more compactly as

$$0 = I_x u + I_y v + I_t \tag{3.6}$$

In order to avoid a confusion between the three-dimensional velocity of a point in space and the two-dimensional velocity of a point on the image plane we will borrow from the terminology of stereo vision and call motion on the image plane *disparity*. Use of the term disparity also emphasizes that although the gradient-based approach is based upon a continuous image function, the technique will always be performed on imagery which is discretely sampled in time and space.

3.2. Gradient Based Algorithms

The gradient constraint equation does not by itself provide a means for calculating disparity. The equation only constrains the values of u and v to lie on a line in disparity space.

The gradient constraint is usually coupled with an assumption that nearby points move in a like manner to arrive at algorithms which solve for disparity. Groups of constrain equations are used to collectively constrain the disparity at a pixel. Constraint lines are combined in one of three ways. The *clustering* approach [1,3] operates globally, looking for groups of constraint lines with coinciding points of intersection in

disparity space. Methods of *local optimization* [32,33,34,35,36] solve a set of constraint lines from a small neighborhood as a system of linear equations. *Global optimization* [*31,37*] techniques minimize an error function based upon the gradient constraint and an assumption of local smoothness of disparity variations over the entire image.

We will examine the local optimization technique in detail. In later sections of the report we will discuss some implications of our analysis for other approaches.

3.3. Local Optimization

The method of local optimization estimates disparity by solving a group of gradient constraint lines obtained from a small region of the image as a system of linear equations. Two constraint lines are sufficient to arrive at a unique solution for (u,v). More than two equations may be included in the system to reduce the affects of errors in the constraint lines. The solution to the over-determined system may be found by any of a number of methods which seek to find a solution which best agrees with a population of constraint lines.

We will begin by examining errors in two equation systems. The pair of equations which we will solve to estimate disparity at point $\mathbf{p}_i = (x_i, y_i, t_i)$ is

$$\begin{array}{ll} (i) & I_x u + I_y v = -I_t \\ (j) & I_x u + I_y v = -I_t \end{array} \tag{3.7}$$

where the gradients I_x, I_y, and I_t in equations i and j are evaluated at nearby points $\mathbf{p}_i$ and $\mathbf{p}_j$.

The gradients in the system (3.7) are estimated from discrete images and will be inaccurate due to noise in the imaging process and sampling measurement error. Also, the values of (u,v) at $\mathbf{p}_i$ and $\mathbf{p}_j$ are assumed to be the same. The formulation will be incorrect to the extent that disparity differs between the two points. We will examine how gradient estimation error and error resulting from nonconstant disparity leads to errors in the estimated disparity.

3.3.1. Gradient Measurement Error

The estimates of the intensity gradient I_x, I_y, and I_t will be corrupted by errors in the brightness estimates and inaccuracies introduced by sampling the brightness function discretely in time and space. The error in the brightness function is random and results from a variety of sources such as channel noise and quantization of brightness levels. The brightness error is approximately additive and independent among neighboring pixels. The gradient, estimated from changes in the brightness estimates, will contain a component of random error which is distributed like the error in the brightness function. The random component of the gradient error will be additive and independent of the magnitude of the gradient to the extent that the brightness noise is additive.

The brightness function is sampled discretely in time and space and this will introduce a systematic measurement error into the estimate of the gradient. The gradient sampling error depends upon the second and higher derivatives of the brightness function. To demonstrate the relationship between the sampling error in $\hat{I}_x$ and the derivatives of brightness we can expand the brightness function evaluated at $(x+\Delta x, y, t)$ around the point (x, y, t) producing

$$I(x+\Delta x, y, t) = I(x, y, t) + I_x \Delta x + \tfrac{1}{2} I_{xx} \Delta x^2 + h.o.t. \tag{3.8}$$

where I_x, I_{xx} are the partial derivatives of brightness in the x direction evaluated at (x, y, t). Rearranging terms we obtain an estimate for the brightness gradient in the x direction:

$$\hat{I}_x = \frac{I(x+\Delta x, y, t) - I(x, y, t)}{\Delta x} = I_x + \tfrac{1}{2} I_{xx} \Delta x^2 + h.o.t. \tag{3.9}$$

The error in the estimate is

$$\varepsilon_{I_x(sampling)} \approx \tfrac{1}{2} I_{xx} \Delta x^2 \tag{3.10}$$

Likewise, the sampling error in the estimates of I_y and I_t are given by

$$\varepsilon_{I_y(sampling)} \approx \tfrac{1}{2} I_{yy} \Delta y^2 \tag{3.11}$$

$$\varepsilon_{I_t(sampling)} \approx \tfrac{1}{2} I_{tt} \Delta t^2 \tag{3.12}$$

The sampling error for the spatial gradients depends upon the spatial resolution of the camera, Δx and Δy, and the second spatial derivatives of the brightness function, I_{xx}, I_{yy}, which in turn depend upon the reflectance characteristics of the surface in view. The sampling error for the temporal gradient, $\varepsilon_{I_t(sampling)}$, is influenced by the frame rate, Δt, and the higher order derivatives of the brightness function over time.

As the gradient constraint equation relates the temporal gradient to the spatial gradients and disparity, the higher order temporal derivatives of the brightness function are related to the higher order spatial derivatives of brightness and the characteristics of the motion on the image plane. Differentiating the gradient constraint equation with respect to x, y, and t we obtain the following three equations:

$$I_{xx}u + I_x\frac{\partial u}{\partial x} + I_{yx}v + I_y\frac{\partial v}{\partial x} = -I_{tx} \tag{3.13}$$

$$I_{xy}u + I_x\frac{\partial u}{\partial y} + I_{yy}v + I_y\frac{\partial v}{\partial y} = -I_{ty} \tag{3.14}$$

$$I_{xt}u + I_x\frac{\partial u}{\partial t} + I_{yt}v + I_y\frac{\partial v}{\partial t} = -I_{tt} \tag{3.15}$$

Where the second derivatives of the brightness function exist and are continuous, the left hand sides of equations (3.13) and (3.14) can be substituted for I_{xt} and I_{yt} in (3.15). Collecting terms we see that

$$\begin{bmatrix}u & v\end{bmatrix}\begin{bmatrix}I_{xx} & I_{xy}\\ I_{yx} & I_{yy}\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix} + u\cdot\begin{bmatrix}I_x\\ I_y\end{bmatrix}\begin{bmatrix}\frac{\partial u}{\partial x} & \frac{\partial v}{\partial x}\end{bmatrix} + v\cdot\begin{bmatrix}I_x\\ I_y\end{bmatrix}\begin{bmatrix}\frac{\partial u}{\partial y} & \frac{\partial v}{\partial y}\end{bmatrix} + \begin{bmatrix}I_x\\ I_y\end{bmatrix}\begin{bmatrix}\frac{\partial u}{\partial t} & \frac{\partial v}{\partial t}\end{bmatrix} = I_{tt} \tag{3.16}$$

The first term in (3.16) depends upon disparity while the rest of the left hand side depends upon the derivatives of disparity over time and space. If disparity is approximately constant in a small neighborhood and approximately constant over time at each point on the image then

$$\begin{bmatrix}u & v\end{bmatrix}\begin{bmatrix}I_{xx} & I_{xy}\\ I_{yx} & I_{yy}\end{bmatrix}\begin{bmatrix}u\\ v\end{bmatrix} \approx I_{tt} \tag{3.17}$$

In summary, then, the systematic errors in the gradients which make up the coefficients of (3.7) are given by (3.9), (3.10), and (3.11). Under the assumption of constant disparity over time and space, the systematic error in the temporal derivative is

$$\varepsilon_{I_t sampling} \approx \tfrac{1}{2}\Delta t^2 \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} I_{zx} & I_{zy} \\ I_{yz} & I_{yy} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.18}$$

The systematic error in estimating I_t increases as the square of disparity. It is also influenced by the derivatives of disparity and the first and second spatial derivatives of brightness. If disparity is significant (3.18) can become quite large and substantially alter our estimate of I_t.

3.3.2. Nonuniformity in the Disparity Field

The estimation scheme which we have been analyzing has assumed that velocity on the image plane is constant in some small neighborhood. This will be true only for very special surfaces and motions. The estimates we obtain will be in error to the extent that disparity varies in the neighborhood which we cover in our equation set. The true set of equations in (3.7) should actually be

$$\begin{array}{lrcl} (i) & I_x u + I_y v & = & -I_t \\ (j) & I_x (u + \Delta u) + I_y (v + \Delta v) & = & -I_t \end{array} \tag{3.19}$$

The difference between the true solution and our estimate can be treated as an error on the right hand side by distributing the multiplication on the left hand side of (3.19) and moving the terms which contain a change in disparity to the right hand side giving

$$\begin{array}{lrcl} (i) & I_x u + I_y v & = & -I_t \\ (j) & I_x u + I_y v & = & -I_t - (I_x \Delta u + I_y \Delta v) \end{array} \tag{3.20}$$

The error caused by the assumption of constant disparity can be treated just as an extra additive error in the estimate of I_t.

3.3.3. Error Propagation

The accuracy of the estimated values of u and v depends upon the magnitude of the errors in the gradient constraint equations and the propagation characteristics of the system of equations which is solved. In this section we examine the error propagation characteristics of the linear system which is solved to estimate disparity.

If the gradients are known exactly and disparity is constant then

$$\mathbf{G}\mathbf{w} = \mathbf{b} \tag{3.21}$$

where,

$$\mathbf{G} = \begin{bmatrix} I_x & I_y \\ I_x & I_y \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} I_t \\ I_t \end{bmatrix} \tag{3.22}$$

As before, the rows of $\mathbf{G}$ and $\mathbf{b}$ are taken from a point $\mathbf{p}_i$ and its neighbor $\mathbf{p}_j$. The vector $\mathbf{w}$ will be in error to the degree that the gradient measurements are inaccurate and disparity varies between points $\mathbf{p}_i$ and $\mathbf{p}_j$. The previous section showed that the error accrued when u and v are not constant is the same as that which would be obtained if the $\mathbf{b}$ vector is suitably modified as in (3.20). This error will be absorbed on the right hand side of (3.21). Thus, the system which is actually solved is

$$(\mathbf{G} + \mathbf{E})(\mathbf{w} + \delta\mathbf{w}) = \mathbf{b} + \delta\mathbf{b} \tag{3.23}$$

where,

$$\mathbf{E} = \begin{bmatrix} \varepsilon_{I_x} & \varepsilon_{I_y} \\ \varepsilon_{I_x} & \varepsilon_{I_y} \end{bmatrix}, \quad \delta\mathbf{b} = \begin{bmatrix} \varepsilon_t \\ \varepsilon_t - (I_x \Delta u + I_y \Delta v) \end{bmatrix} \text{ and } \delta\mathbf{w} = \begin{bmatrix} \varepsilon_u \\ \varepsilon_v \end{bmatrix} \tag{3.24}$$

Distributing the multiplication of $(\mathbf{G} + \mathbf{E})$ and rearranging terms we see that

$$\mathbf{G}(\mathbf{w} + \delta\mathbf{w}) = \mathbf{b} + \delta\mathbf{b} - \mathbf{E}(\mathbf{w} + \delta\mathbf{w}) \tag{3.25}$$

Consequently,

$$\mathbf{w} + \delta\mathbf{w} = \mathbf{G}^{-1}\left[\mathbf{b} + \delta\mathbf{b} - \mathbf{E}(\mathbf{w} + \delta\mathbf{w})\right] \tag{3.26}$$

Since $\mathbf{w} = \mathbf{G}^{-1}\mathbf{b}$, we have

$$\delta\mathbf{w} = \mathbf{G}^{-1}\left[\delta\mathbf{b} - \mathbf{E}(\mathbf{w} + \delta\mathbf{w})\right] \tag{3.27}$$

The vector $\delta\mathbf{w}$ is the absolute error in the disparity estimate. The absolute error depends upon the inverse of the matrix of spatial gradients, the error vectors, and the disparity vector itself.

Let us divide $\delta\mathbf{w}$ into two components

$$\delta\mathbf{w} = \delta\mathbf{w}_{lhs} - \delta\mathbf{w}_{rhs} \tag{3.28}$$

where,

$$\delta\mathbf{w}_{lhs} = \mathbf{G}^{-1}\mathbf{E}(\mathbf{w} + \delta\mathbf{w}) \quad \text{and} \quad \delta\mathbf{w}_{rhs} = \mathbf{G}^{-1}\delta\mathbf{b} \tag{3.29}$$

This decomposition separates $\delta\mathbf{w}$ into two components which depend principally upon errors in the left and right hand sides of (3.21). To see how large these errors might be we take the norms of $\delta\mathbf{w}_{rhs}$ and $\delta\mathbf{w}_{lhs}$ and find that

$$\|\delta\mathbf{w}_{lhs}\| \leq \|\mathbf{G}^{-1}\| \cdot \|\mathbf{E}\| \cdot \|(\mathbf{w} + \delta\mathbf{w})\| \tag{3.30}$$

and

$$\|\delta\mathbf{w}_{rhs}\| \leq \|\mathbf{G}^{-1}\| \cdot \|\delta\mathbf{b}\| \tag{3.31}$$

It follows directly from (3.30) that

$$\frac{\|\delta\mathbf{w}_{lhs}\|}{\|(\mathbf{w} + \delta\mathbf{w})\|} \leq \|\mathbf{G}^{-1}\| \cdot \|\mathbf{E}\| = cond(\mathbf{G}) \frac{\|\mathbf{E}\|}{\|\mathbf{G}\|} \tag{3.32}$$

Where, the *condition number*, represented by $cond(\mathbf{G})$, is defined to be the value of $\|\mathbf{G}\| \cdot \|\mathbf{G}^{-1}\|$ for any nonsingular matrix $\mathbf{G}$ [38]. Since $\mathbf{Gw} = \mathbf{b}$, we know that

$$\|\mathbf{w}\| \geq \frac{\|\mathbf{b}\|}{\|\mathbf{G}\|} \tag{3.33}$$

Dividing (3.31) by (3.33):

$$\frac{\|\delta\mathbf{w}_{rhs}\|}{\|\mathbf{w}\|} \leq \|\mathbf{G}\| \cdot \|\mathbf{G}^{-1}\| \cdot \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} = cond(\mathbf{G}) \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{3.34}$$

The relative error in both components depends upon the condition number of the matrix of spatial gradients. In turn, the conditioning of the $\mathbf{G}$ matrix is determined by the nature of the brightness function over the interval $(\mathbf{p}_i, \mathbf{p}_j)$. We can express the spatial gradients at $\mathbf{p}_j$ as a function in terms of the spatial gradients and higher order derivatives of the brightness function at the point $\mathbf{p}_i$. Expanding the brightness gradient evaluated at $\mathbf{p}_j$ in a Taylor's series around the point $\mathbf{p}_i$ we obtain

$$I_x(x_j,y_j,t) = I_x(x_i,y_i,t) + I_{xx}(x_i,y_i,t)\Delta x + I_{xy}(x_i,y_i,t)\Delta y + h.o.t. \tag{3.35}$$

and

$$I_y(x_j,y_j,t) = I_y(x_i,y_i,t) + I_{xy}(x_i,y_i,t)\Delta y + I_{yy}(x_i,y_i,t)\Delta y + h.o.t. \tag{3.36}$$

where,

$$\Delta x = x_i - x_j \quad \text{and} \quad \Delta y = y_i - y_j$$

Approximating the gradient at $\mathbf{p}_j$ by the first three terms of the Taylor's series expansions in (3.35) and (3.36) we arrive at an expression for the matrix of spatial derivatives:

$$\mathbf{G} \approx \begin{bmatrix} I_x & I_y \\ I_x + I_{xx}\Delta x + I_{xy}\Delta y & I_y + I_{yx}\Delta x + I_{yy}\Delta y \end{bmatrix} \tag{3.37}$$

All of the terms in (3.37) are evaluated at $\mathbf{p}_i$. Without loss of generality, we can rotate the spatial coordinates so that at the point $\mathbf{p}_i$ in the new coordinate system (x,y,t),

$$I_{xy} = I_{yx} = 0$$

The simplified matrix of spatial derivatives is

$$\mathbf{G} = \begin{bmatrix} I_x & I_y \\ I_x + I_{xx}\Delta x & I_y + I_{yy}\Delta y \end{bmatrix} \tag{3.38}$$

The inverse of (3.38) is easily calculated as

$$\mathbf{G}^{-1} = \frac{1}{I_x I_{yy}\Delta y - I_y I_{xx}\Delta x} \begin{bmatrix} I_y + I_{yy}\Delta y & -I_y \\ I_x + I_{xx}\Delta x & I_x \end{bmatrix} \tag{3.39}$$

The magnitude of $\|\mathbf{G}^{-1}\|$ depends upon the first and second derivatives of brightness and the relative position of the two points at which the brightness function is evaluated. The second derivatives clearly must be nonzero or $\mathbf{G}$ will be singular and $\mathbf{G}^{-1}$ will not exist.

Measurement errors in the brightness gradients are multiplied by $\mathbf{G}^{-1}$ to determine to absolute error in the disparity estimate. It would seem that the disparity estimate would be most accurate when $\|\mathbf{G}^{-1}\|$ is smallest. This is true for the random component of measurement error and the portion of error due to nonconstant disparity. However, the opposite may be true for the systematic measurement errors.

The propagation of systematic measurement error is complicated by the fact that both $\mathbf{G}^{-1}$ and the systematic measurement errors in the gradients depend upon the nature of the second derivatives of brightness. The systematic errors in $\hat{I}_x$, $\hat{I}_y$, and $\hat{I}_t$ increase with the magnitudes of the higher order derivatives of brightness. The magnitude of $\|\mathbf{G}^{-1}\|$ also depends, in part, upon the magnitude of the second (and higher order) derivatives of brightness. For a given brightness function, we are free to choose the direction and distance of the neighboring point which contributes the second equation to the linear system. This determines the difference vector $(\Delta x, \Delta y)$ in (3.39). Let us fix the distance to the neighbor and assume that the orientation of the difference vector is chosen so as to minimize the norm of (3.39). Under these circumstances, increases in the second derivative can lead to a reduction in the magnitude of $\|\mathbf{G}^{-1}\|$.

The systematic measurement error and the random measurement error are oppositely affected by variations in the linearity of the brightness function. If the spatial gradients are nearly constant then random measurement errors and the error due to non-constant disparity will be greatly magnified in the solution vector. If, however, the spatial gradients rapidly vary then the solution vector may be overly corrupted by the systematic measurement error. Accurate estimates can be reached only when both sources of error are relatively small.

The systematic error in the temporal gradient increases as the square of the magnitude of disparity. So, if disparity is constant then the systematic error in the temporal gradient is negligible and most accurate estimates will be obtained when the brightness function is very nonlinear. In regions where disparity is large the systematic error in the temporal gradient will be very sensitive to nonlinearities in the brightness function and the best estimates will be obtained when the brightness function is approximately linear over the region of translation.

The propagation characteristics of **G** can be improved by increasing the distance, d, to the neighbor which contributes the second constraint equation. The risk in choosing neighbors over too great a distance is that the error due to nonconstant disparity will become very large. Disparity will tend to vary smoothly across object surfaces. If neighbors lie on a different surfaces their motions may differ substantially. As the distance to the neighbor is increased it becomes more likely that the difference in disparity between neighbors will contribute a significant error to the system.

The error in the estimate of disparity is determined by the characteristics of the disparity field, the nature of the brightness function, and the selection of rule for constructing the linear system. These parameters interact in a complex way to determine the accuracy of the local optimization scheme. More study is required to better understand the this interaction before precise performance bounds can be obtained for the technique.

4. ALGORITHM EXTENSIONS BASED UPON ERROR ANALYSIS

The previous section identified the major determinants of error for disparity estimated from a gradient-based method. In this section knowledge about the causes of errors is used to explain how errors can be reduced and to introduce techniques to judge the accuracy of estimates. The improvements in performance are based upon parameter selection and preprocessing of the image to extract the most information from a region while minimizing the intrusions of error. A method of iterative refinement [35] is also described.

By examining the image sequence for the conditions which lead to errors we can judge the accuracy with which estimates can be made before the estimate is actually made. Examination of the disparity estimate itself can provide additional information about the precision of the estimate. Together, *a priori* and *a posteriori* estimates of accuracy provide a useful heuristic for evaluating the precision of disparity estimates.

4.1. Error Reduction Techniques

Several techniques can be used to improve the accuracy of the disparity estimates obtained with the local optimization technique. Blurring the image will reduce nonlinearities in the brightness funciton and consequently diminish the systematic error in the gradient estimates. Blurring will also worsen the propagation characteristics of the linear system causing random measurement errors and the errors due to nonconstant disparity to be magnified. Hence, blurring is desirable only in regions where the systematic error is predominant.

As noted in the last section, the systematic error in the gradients depends upon the nonlinearity of the brightness function over the sampling interval. For the temporal gradient, the systematic measurement error depends upon the linearity of the brightness function over the region of motion and the variations of disparity over time and space. Blurring will be most effective in portions of the image which undergo a

translation through a region of nonlinearity. The degree of blurring should be sufficient to linearize the brightness function over the region of translation.

The damage which blurring does to the propagation characteristics of the linear system can be counterbalanced by increasing the size of the neighborhood over which the system is constructed. The risk incurred by enlarging the area from which the constraint equations are drawn is that the motions of the points may differ significantly, as could happen if points lied on two different surfaces. The selection of the radius of blur and the neighborhood size must be made judiciously so as to avoid increasing the error in the solution vector.

Until this point we have ignored the problem of selecting the direction in which the neighbor is to be chosen to form the linear system. From our previous discussion of error propagation it is clear that the choice of direction can dramatically affect the error in the disparity estimate. One way to circumvent the difficulty of choosing an appropriate direction is to construct an over-determined set of equations from points in many directions. The over-determined system can be solved by minimizing the residual over possible values of disparity. The conditioning of the over-determined system is about the same as the conditioning a system based upon the optimal pair of equations in the set. The norm which is minimized may have an important affect on the sensitivity of the system to some kinds of errors. More study is required to determine the influence of the minimization criteria on the accuracy of solutions. Another approach is to perform the analysis separately in a number of directions and then seek a consensus among solutions [39]. If the errors are random then the estimates will tend to be distributed about the true value of disparity. Both approaches have the advantage of extracting the important information about motion from a region without explicitly searching for where the information is concentrated.

If disparity is known approximately then this knowledge can be used to reduce the error in the local optimization technique. Let $\hat{\mathbf{w}}$ be a three-dimensional vector which

describes the velocity of an object point through the three-dimensional image function $I(x,y,t)$. Let $\mathbf{w}$ be the true disparity, and $\delta\mathbf{w}$ be the difference between the true disparity and the estimate. It follows from our definitions that

$$\hat{\mathbf{w}} + \delta\mathbf{w} = \mathbf{w} \tag{4.1}$$

where,

$$\mathbf{w} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad \hat{\mathbf{w}} = \begin{bmatrix} \hat{u} \\ \hat{v} \\ 1 \end{bmatrix} \quad \text{and} \quad \delta\mathbf{w} = \begin{bmatrix} \delta u \\ \delta v \\ 1 \end{bmatrix} \tag{4.2}$$

The derivative of brightness in the direction of the estimated disparity is

$$I_{\hat{\mathbf{w}}} = I_x\hat{u} + I_y\hat{v} + I_t \tag{4.3}$$

Using the gradient constraint equation and substituting for I_t we see that

$$I_{\hat{\mathbf{w}}} = I_x\hat{u} + I_y\hat{v} - I_x u - I_y v \tag{4.4}$$

and consequently,

$$0 = I_x\delta u + I_y\delta v + I_{\hat{\mathbf{w}}} \tag{4.5}$$

Equation (4.5) is a more general form of the gradient constraint equation which relates the gradient in an arbitrary direction to the spatial gradients and disparity The derivative in the direction of the motion estimate can be approximated by

$$I_{\hat{\mathbf{w}}} = I(x+\hat{u},y+\hat{v},t+\delta t) - I(x,y,t) \tag{4.6}$$

If the disparity estimate is (0,0,1) then $I_{\hat{\mathbf{w}}}=I_t$ and we obtain the familiar gradient constraint equation. All of the analysis performed thus far applies to the more general form of the gradient constraint equation.

We can use the general form of the gradient constraint equation to refine an estimate $\hat{\mathbf{w}}$ by solving for $\delta\mathbf{w}$. This process can be performed iteratively to find successively better estimates of disparity. An improvement can be expected, on the average, whenever successive estimates are closer to the true disparity.

$$\| \delta \mathbf{w}_{i+1} \| \leq \| \delta \mathbf{w}_i \| \qquad i = 1, 2, \ldots \tag{4.7}$$

The improvement arises from successively better estimates of the directional derivative $I_{\underline{\mathbf{w}}}$. As was demonstrated earlier in equation (3.16) the systematic error in the estimate of temporal derivative grows as the square of disparity. The same relationship is true for direction derivative $I_{\underline{\mathbf{w}}}$ and the disparity difference in the general constraint equation.

Solving for the difference between an estimate of disparity and the true disparity is computationally equivalent to registering a portion of an image pair and estimating the change of position in the adjusted sequence For this reason the technique has been called *iterative registration* [35]. The estimate of disparity may be derived from estimates made at some previous time or from prior processing on a single frame pair.

Note that if the inequality of (4.7) does not hold then the error might be expected to increase. If an estimate of disparity is poor then the refinement effort may lead to an even larger error. The next section is devoted to methods to evaluate the quality of disparity estimates. A measure of the accuracy of a disparity estimate can be used to judge whether or not the estimate should be used for registration. Alternatively, the degree of registration can be based upon the confidence which can be put in the disparity estimate, the more accurate the estimate is judged to be, the more that the frame pair should be adjusted in the direction of the estimate.

The iterative registration technique can be combined with variable blurring to produce a coarse-to-fine system for estimating disparity [35]. Disparity is roughly estimated with an image sequence which has been blurred sufficiently to linearize the brightness function over the maximum expected displacement. The coarse estimate of disparity is used, at each point, to register a small region of the image at a finer level of resolution. This process is repeated at successively finer levels of resolution.

How much advantage can be gained from iterative registration? Motion differentials will be the same for all registrations. Thus, the error due to incompatibilities among equations in the linear system is unaffected by iterative registration. Also, the estimate of the directional gradient will contain some amount of random measurement error even if successive frames are in perfect registration. The propagation of these errors depends primarily upon the magnitude of $\|\mathbf{G}^{-1}\|$. We can not expect to reduce the error in $\hat{\mathbf{w}}$ below that caused by random error in $I_{\hat{\mathbf{w}}}$ and nonconstant disparity through iterative registration.

While preforming a coarse-to-fine registration the degree of blurring at each stage should be appropriate to the expected error in disparity estimated at the next more coarse level of analysis. In the absence of knowledge about the motions of individual points the blurring must be performed uniformly across the image. While the error will, on the average, be reduced for points which translate significantly, the error will tend to be increased for points which are stationary or move very little. No benefit is obtained by linearizing the brightness function at stationary regions and the error propagation characteristics are worsened. Some of the accuracy lost at stationary regions during coarse processing might be recovered at finer levels but, in general, the best estimates could be obtained at a fine level without registration. In the next section methods are developed to estimate the accuracy of disparity estimates. This information can be used in the coarse-to-fine system of iterative registration to judge whether an improvement has been obtained at each level. A priori estimates of the magnitude of disparity are also developed in the next section. The iterative registration technique can be improved by adapting the technique to knowledge about the accuracy of estimates and the magnitude of motion.

4.2. Estimating Error

Many of the factors which lead to errors in the local optimization estimation technique can be identified and measured from the image. The error propagation characteristics of the linear system $\mathbf{Gw} = \mathbf{b}$ can be estimated from the matrix of spatial gradients. Random errors in the estimates of the gradients and errors due to variations in the disparity field are magnified by $\|\mathbf{G}^{-1}\|$ in the solution vector. The degree to which relative errors are magnified is indicated by $cond(\mathbf{G})$. Regions of the image for which the propagation characteristics are poor will be very sensitive to small measurement errors in the gradients. The disparity estimates obtained in these regions are likely to be inaccurate.

The systematic measurement error in $\hat{I}_t$ was shown to depend upon the linearity of the brightness function over the interval of translation. One way to measure of the nonlinearity of the brightness function, suggested by [35] and [3], is to compare the spatial gradients of brightness in successive frames. If $I_x(x,y,t)$ is significantly different from $I_x(x,y,t+\delta t)$ then it can be inferred that the estimate of the temporal gradient is likely to be in error.

The magnitude of the disparity vector is also an important determinant of the systematic measurement error in $\hat{I}_t$. The error in the estimate of the temporal gradient grows as the square of the disparity. The size of the disparity vector can be bounded by examining the brightness gradients. The gradient constraint equation can be written as

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \tag{4.8}$$

The size of the disparity vector can be bounded by

$$\left\| \begin{bmatrix} u \\ v \end{bmatrix} \right\|_2 \geq \frac{|-I_t|}{\left\| \begin{bmatrix} I_x & I_y \end{bmatrix} \right\|_2} = \left| \frac{I_t}{\sqrt{I_x^2 + I_y^2}} \right| \tag{4.9}$$

The relationship between the temporal and spatial gradients in (4.9) is a useful heuris-

tic for judging the accuracy of $\hat{I}_t$. By performing the same manipulation on the generalized constraint equation (4.5), we also arrive at a means to evaluate the accuracy of our disparity estimate. If, as before, $I_{\hat{\mathbf{w}}}$ is the derivative of brightness in the direction of the estimated disparity then

$$\left\| \begin{bmatrix} \delta u \\ \delta v \end{bmatrix} \right\|_2 \geq \frac{|-I_{\hat{\mathbf{w}}}|}{\left\| \begin{bmatrix} I_x & I_y \end{bmatrix} \right\|_2} = \left| \frac{I_{\hat{\mathbf{w}}}}{\sqrt{I_x^2 + I_y^2}} \right| \tag{4.10}$$

Thus, we can estimate the derivative in the direction of the disparity estimate by comparing the brightness function at point (x,y,t) to the brightness function at time $t+\delta t$ evaluated at the estimated translation $x+\hat{u}$, $y+\hat{v}$. This difference, without consideration of the magnitudes of the spatial gradients, has been called the *displaced frame difference* and is an important component in a scheme for coding television signals [33]. If the norm of the spatial gradients is not too small, the displaced frame difference divided by the magnitude of the spatial gradients is an good measure of the magnitude of the error in the disparity estimate.

If an overdetermined set of equations is used to estimate disparity then measurement errors in the gradients and incompatibilities among the constraint equations due to differential motion will be reflected in the residual of the solution. The residual vector can be estimated by

$$\mathbf{G}\hat{\mathbf{w}} - \mathbf{b} = \mathbf{r} \tag{4.11}$$

where $\hat{\mathbf{w}}$ is the estimated disparity and $\mathbf{r}$ is the residual. Conversely, A large residual indicates that substantial errors exist in the system and that the estimated disparity vector is likely to be inaccurate.

The residual vector will be especially large at occlusion edges where the change in disparity is discontinuous. It has been proposed [34] that the residual error be used as an indication of the presence of an occlusion edge. To be identifiable, the change in disparity across an occlusion edge must lead to an error which is greater then that

normally encountered from other measurement errors. A threshold on the residual must be established which will normally be exceeded only at significant discontinuities in the disparity field. The error accrued from a change in disparity is equivalent to a measurement error on the right hand side of the local optimization system. Since the equivalent error on the right hand side is magnified by the size of the spatial gradients (3.29), the threshold for identifying large residual errors may be adaptive to the spatial gradients. Likewise, it was shown that the systematic measurement errors in the gradients were related to the second derivatives of brightness, so the threshold on the residual may depend upon the second derivatives, as well.

4.3. Summary

The gradient constraint is a powerful tool for the analysis of dynamic imagery. Careful examination of one gradient-based technique has led to a number of conclusions about the causes of errors, provided support for techniques to improve estimates, and indicated methods by which the accuracy of estimates could be judged. This analysis suggests that disparity estimation should be adaptive to the nature of the brightness function and the characteristics of motion in a region of the image. Empirical investigations support the analytical work presented here. More research is needed to understand the magnitude of the error bounds presented here and elaborate on the analysis.

5. Empirical Analysis

5.1. Methods

5.1.1. Feature Point Selection

The two matching techniques which are described below attempt to find correspondences only in regions of high information content. A variety of methods for identifying *feature* or *interest points* have been proposed, including adaptive templates [6,7], local maxima of variability [40,41,24,13], local extrema of the laplacian [42,27], maxima of gaussian curvature [43], and local edges [44]. Although no consensus on the best approach exists, it is clear that feature points should be efficiently computable and should reliably locate the same points on objects from frame to frame. It is important not only that the population of feature points represent the same object features from frame to frame, but also that the position of the feature point on the object should be stable. For example, most methods for locating feature points favor corners in the brightness function. If a corner is identified as a region of interest in one frame it is desirable that it be selected in subsequent frames as well. Further, it is desirable that the placement of the feature point on the corner always occur in the same location.

The distribution criteria developed to evaluate disparity fields (section 2.2.2) also apply to the feature point selection process, because correspondences can only be obtained at feature points. As such, algorithms which select feature points can be judged by the density and dispersion of the feature points.

For the matching techniques demonstrated here, feature points were selected as local extrema in the laplacian. An approximation of the laplacian is efficiently computed by the subtracting two versions of an image which have been differentially blurred [45,46]. Choosing the maxima and minima of the laplacian produces a rela-

tively dense sampling of feature points which are well dispersed across the image. These feature points tend to associated with distinctive structures in the image. An additional characteristic of this technique is that the size of the structures which are associated with feature points can be affected by the amount of blurring to which the images are subjected. This aspect of the laplacian method makes it easily adapted to coarse-to-fine techniques.

5.1.2. Relaxation Matching

The relaxation matching algorithm is shown schematically in figure 5.1. The matching process starts with determination of all possible correspondences between feature points in the two frames. A bound of 15 pixels was placed on the maximum possible disparity for the sequences used in our examples. A list of possible matches in the second frame is created for each feature point in the first frame. The list of candidate matches is structured as a set of labels. In addition, a unique label l^* is added to the list. The l^* label represents the condition that there is no correct match in the second frame. Thus, each feature point f^i in the first frame has an associated label list of the for

$$L^i=\left\{l_1^i, l_2^i, \ldots, l_m^i, l^*\right\} \quad where, \; l_j^i = (u,v)_j^i \tag{5.1}$$

The matching task is to choose the correct label for every feature point, f^i. A confidence value p_j^i is used as an estimate of the likelihood that label j is a correct match for point i. Initial estimates for p_j^i are found by correlating a 5×5 window around the point f^i with a similar window around each of the potential matches in the second frame. The initial estimate for l^* is based upon the magnitude of the correlation values found for other possible matches: if none of the possible matches in second frame correlate well with f^i, then l^* is, initially, given a large value.

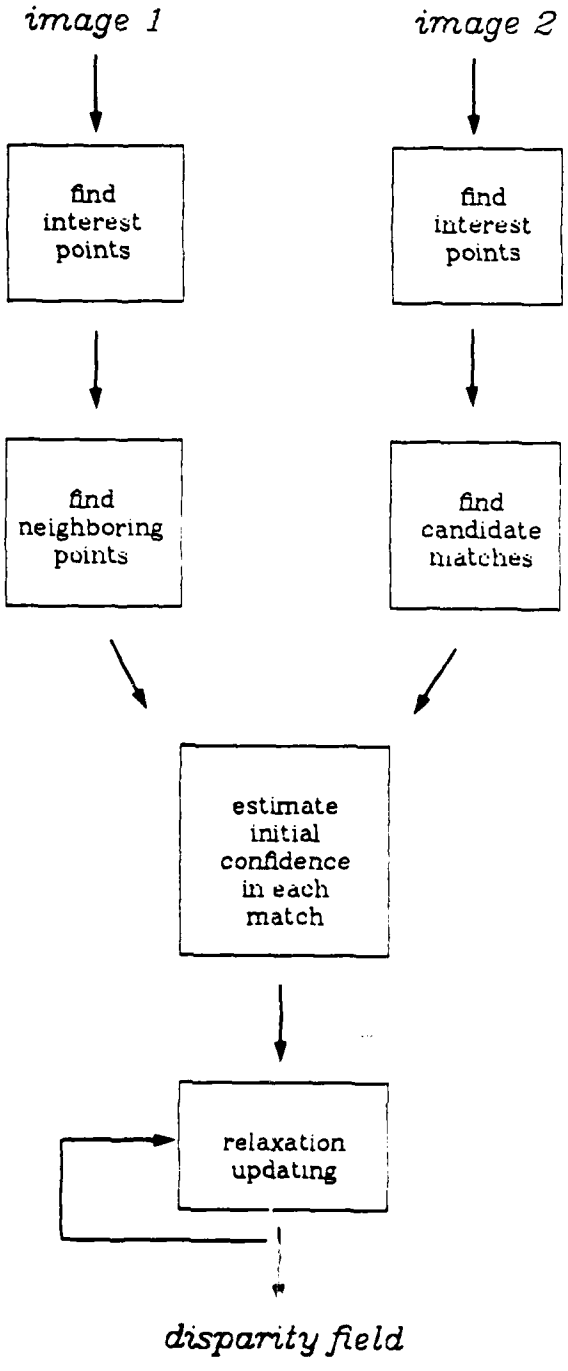


Figure 5.1 Token Matching with relaxation labeling

The initial confidence estimates are refined through an iterative process that incorporates constraints about the possiole disparities for each point. If surfaces are large relative to the sampling distribution, disparity will vary slowly over most of the image. Thus, nearby points will usually move in a similar way. Estimates are refined by comparing the list of possible matches for neighboring points and favoring consistencies in motion. A sequence of progressively more precise estimates of label likelihoods $L^i(1), L^i(2), \ldots$ is generated by letting

$$p_j^i(k+1) = N\left[p_j^i(k)\cdot(1+\alpha(q_j^i(k))\right] \tag{5.2}$$

on iteration $k+1$, where,

$q_j^i(0) \equiv$ the initial estimates described above,

$q_j^i \equiv$ a function that measures the number and likelihood of labels in the neighborhood of f^i of the same or nearly the same disparity as l_j^i,

$\alpha \equiv$ an adjustable gain parameter, and

$N \equiv$ a normalization function that assures the sum of likelihoods over a given label set to be always equal to 1.

The q function assigns a large value to l^* if candidate matches frequently occur in neighbors' lists at a disparity which does not exist in the label set L^i. The effect of (5.2) is to raise the the likelihood of possible matches if there are other, nearby, high-likelihood matches with the same disparity. By iterating the estimation process, information can propagate through the network of feature points. Some points are easily matched on the basis of correlation alone. This affects the label likelihoods for nearby points, and, as the estimates improve for those points, they, in turn influence their neighbors. The process converges if and when the likelihood of one label dominates all the others in each label set. A more detailed presentation of the relaxation method can be found in [47].

5.1.3. Correlation Matching

The correlation-based matching program implements a method developed by Moravec [40,41,24]. Feature points are located in the first of two sequentially taken images. The second image is searched to find the position which best matches the feature point in the first image. A small window centered around a feature point in the first image, the *feature window*, is compared to similar sized windows in the second image. The search is restricted to windows which lie within a larger *search window*. In our work the feature window was 5×5 and the search window was 7×7.

The analysis proceeds in a coarse-to-fine manner. The image pair is reduced by successively halving the sampling rate. In our examples the initial 128×128 image is reduced three times to give four levels of resolution:

	image reduction
coarsest level	16 × 16
	32 × 32
	64 × 64
finest level	128 × 128

Table 5.1

The search begins at the coarsest level of resolution. A search window is centered around the location of the feature point in the first image and the best match is determined. A pseudo-normalized measure of the cross correlation [40] is used as the criteria function.

The disparity estimate obtained at the coarsest level is used to center the search window at the next finer level of resolution. The process in repeated at each finer level of resolution, always centering the search window on the estimate from the previous level. The method is schematically outlined in figure 5.2.

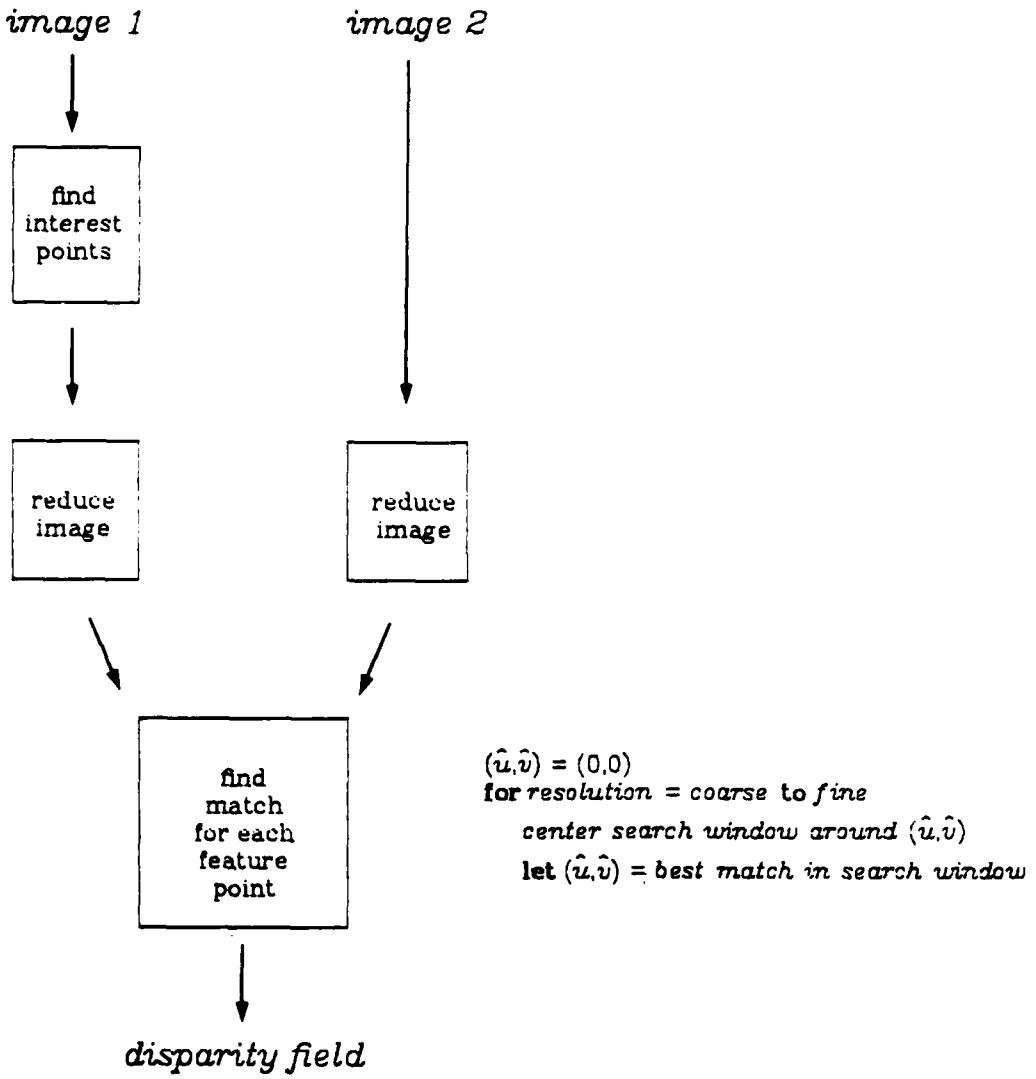


Figure 5.2 Token Matching with coarse-to-fine correlation.

5.1.4. Gradient-Based Estimation

The gradient-based approach is demonstrated with two versions of the local optimization technique. In the next section, a variation of the global optimization technique is introduced as means of combining matching and gradient-based techniques.

5.1.4.1. The Flow Field Data Structure

A scheme for representing the optical flow field is necessary. In our work the vector field is represented as dense image which is spatially registered with the gray-level images of the scene. Each pixel in the disparity image is a three-dimensional vector containing

u - disparity in the x direction

v - disparity in the y direction

p - confidence

The confidence value can range from 0.0 to 1.0.

5.1.4.2. Simple Local Optimization

The basic local optimization method performs a least squared minimization on an over-determined set of gradient constraint equations to estimate disparity at each point. The system is schematically shown in figure 5.3. Each image is first blurred with a gaussian blurring function. The standard deviation of the blurring function used to collect the data presented here was approximately 2 pixels. The blurring serves to reduce the noise in the image and linearize the brightness function.

In the next stage of processing the gradient constraint equations are determined. Gradients are estimated by the difference in the blurred brightness estimates $I(x,y,t)$. The gradient computations are graphically shown in figure 5.4. The gradients at a point (i,j) on the image are estimated as follows,

$$I_x(i,j,t) = \frac{1}{8}\Big[I(i+1,j,t) - I(i-1,j,t) + I(i+1,j,t+1) - I(i-1,j,t+1)\Big] \tag{5.3}$$

$$I_y(i,j,t) = \frac{1}{8}\Big[I(i,j+1,t) - I(i,j-1,t) + I(i,j+1,t+1) - I(i,j-1,t+1)\Big] \tag{5.4}$$

$$I_t(i,j,t) = \frac{1}{2}\Big[I(i,j,t+1) - I(i,j,t)\Big] \tag{5.5}$$

With this method, the gradients estimates are spatially registered with the image pair and temporally sequenced at time between the two frames.

Constraint equations from a group of neighboring points are gathered to produce an over-determined system of linear equations of the form

$$\mathbf{Gw} = \mathbf{b} \tag{5.6}$$

where,

$$\mathbf{G} = \begin{bmatrix} I_x & I_y \\ I_x & I_y \\ \vdots & \vdots \\ I_x & I_y \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} u \\ v \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} I_t \\ I_t \\ \vdots \\ I_t \end{bmatrix} \tag{5.7}$$

The rows of **G** and **b**, are taken from a point (i,j) and a group of nearby points selected from the neighborhood $(i-n,j-n),\ldots,(i+n,j+n)$. To insure that the equations are sufficiently distinct we selected neighbors from a 5×5 window centered around the point to be estimated. The distribution of constraint equations is diagrammed in figure 5.5.

	i-2		i		i+2	
j-2	*N*		*N*		*N*	
j	*N*		*p*		*N*	
j+2	*N*		*N*		*N*	

Figure 5.5. The linear system (5.6) is constructed from the constraint equations evaluated at the point p and its neighbors N.

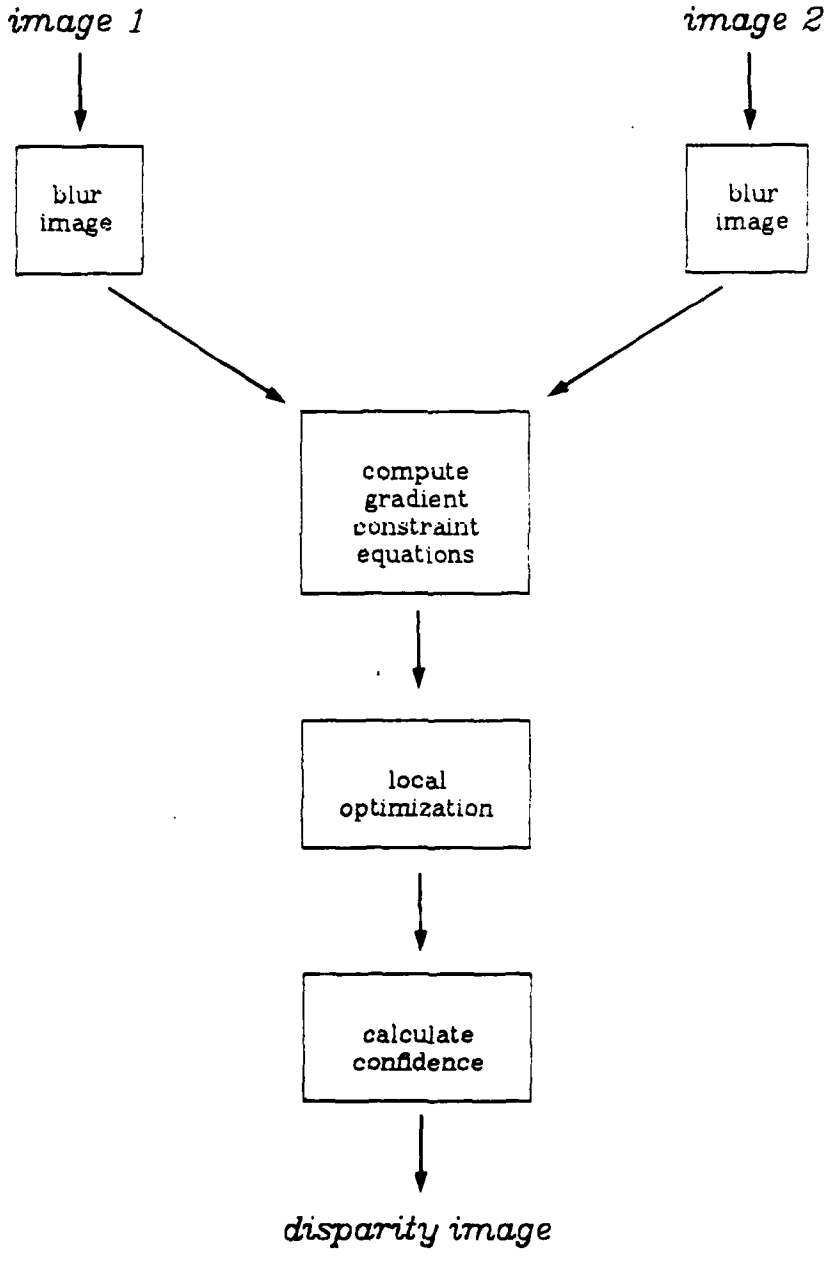


Figure 5.3 The simple local optimization technique.

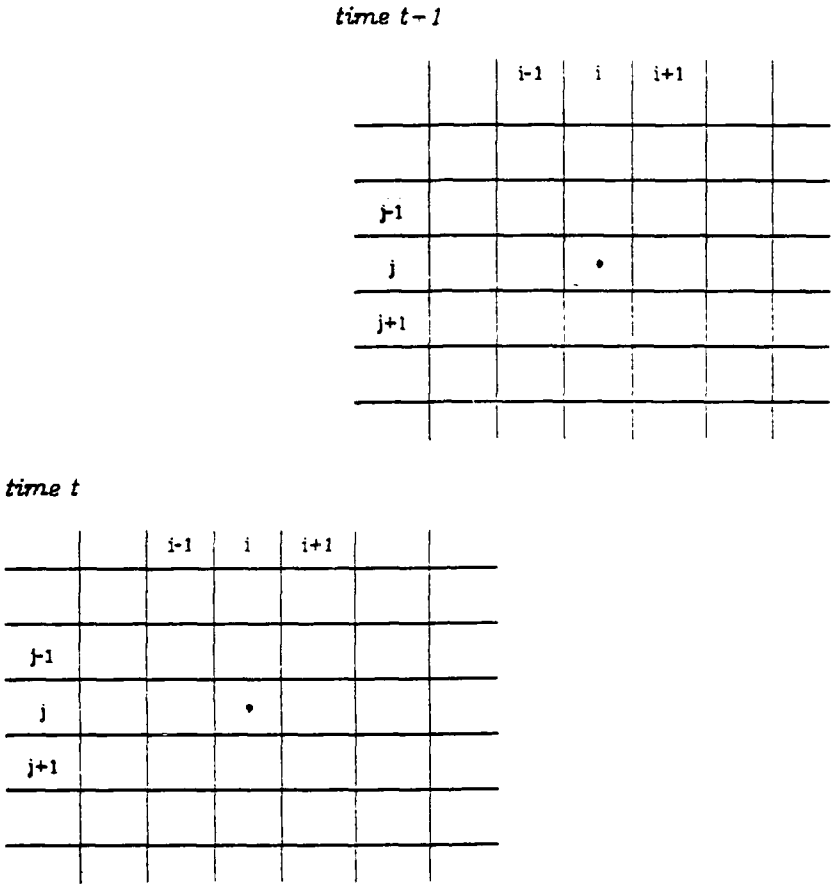


figure 5.4 The gradients are estimated from the blurred image brightness function.

In general, the over determined system (5.6) has no exact solution. An approximate solution is found by minimizing the residual vector $\mathbf{r}$, defined as

$$\mathbf{G}\hat{\mathbf{w}} - \mathbf{b} = \mathbf{r} \tag{5.8}$$

The disparity estimate is chosen to be the vector $\mathbf{w}$ which minimizes some criteria function of $\mathbf{r}$. In our work we minimize $\|\mathbf{r}\|_2$ by letting

$$\hat{\mathbf{w}} = \mathbf{G}^{+}\mathbf{b} \tag{5.9}$$

where $\mathbf{G}^{+}$ is the pseudoinverse of $\mathbf{G}$ [34]. The pseudoinverse is calculated as

$$\mathbf{G}^{+} = \left[\mathbf{G}^{t}\mathbf{G}\right]^{-1}\mathbf{G}^{t} \tag{5.10}$$

This computation requires the inversion of the 2×2 matrix $\mathbf{G}^t\mathbf{G}$. The inverse will not exist where the local gradients do not sufficiently constrain disparity to allow for an exact solution. In this case the confidence of the disparity estimate is set to zero and u and v are undefined.

A confidence is assigned to each disparity estimate on the basis of:

(a) the size of the residual vector $\mathbf{r}$,

(b) the change in the spatial gradients over the temporal sampling interval,

(c) the brightness difference for the image pair registered by (u,v), and

(d) the magnitude of $\|G^{-1}\|$.

The importance of each of these factors in determining the accuracy of estimates is discussed in section 4.2. That analysis does not, however, provide us with a formula for estimating the total error in the disparity vector (u,v). We must find a means to combine several factors which each indicate the presence of conditions which can be lead to errors.

Recall how each factor outlined above relates to the error in (u,v). The residual vector indicates the degree to which the estimated disparity vector jointly satisfies the system of constraint equations. The units associated with the residual vector are not easily interpreted. To obtain a measure of the expected error in pixel units we determine the average minimum distance from $(\hat{u},\hat{v})$ to the gradient constrain equations that make up the system. The minimum distance from $(\hat{u},\hat{v})$ to the gradient constraint line $I_x u + I_y v + I_t$ is easily computed as

$$d = \frac{I_x\hat{u} + I_y\hat{v} + I_t}{I_x^2 + I_y^2} \tag{5.11}$$

The confidence in the estimate, based upon the residual, is

$$p_1 = \frac{1}{\bar{d}^2+1} \tag{5.12}$$

where $\bar{d}$ is the average minimum distance between $(\hat{u},\hat{v})$ and the constraint equations

that make up the linear system. If p_1 is small, then the disparity estimate does not well satisfy the mutual constraints from nearby points and is likely to be in error.

Measurement error in the temporal gradient depends upon the linearity of the brightness function over the translation interval. The change in the spatial gradients between successive frames provides an indication of the linearity of the brightness function over the region which has translated by a point [35]. The variation in the spatial gradients primarily contributes to measurement error in I_t, which lies on the right hand side of the gradient constraint equation. To obtain an estimate of the magnitude of error in $\hat{\mathbf{w}}$ we must divide errors on the right hand side by the magnitude of the spatial gradient. Thus, we estimate the error caused by nonlinearity in the brightness function by the ratio of the change in the spatial gradients to the magnitude of the spatial gradient:

$$\hat{\varepsilon}_\Delta = \frac{\sqrt{\Delta I_x^2 + \Delta I_y^2}}{\sqrt{I_x^2 + I_y^2}} \tag{5.13}$$

where the spatial gradients are estimated as in (5.3) and (5.4) and the changes in the spatial gradients are estimated by

$$\Delta I_x = \tfrac{1}{2}\Big[I(i+1,j,t)-I(i-1,j,t)-I(i+1,j,t+1)+I(i-1,j,t+1)\Big] \tag{5.14}$$

$$\Delta I_y = \tfrac{1}{2}\Big[I(i,j+1,t)-I(i,j-1,t)-I(i,j+1,t+1)+I(i,j-1,t+1)\Big] \tag{5.15}$$

The inverse of the error estimate is used to estimate the confidence in $(\hat{u},\hat{v})$.

$$p_2 = \frac{1}{\hat{\varepsilon}_\Delta + 1} \tag{5.16}$$

The confidence value p_2 gives a rough estimate of the likelihood the $(\hat{u},\hat{v})$ is in error due to measurement error in the temporal gradient.

One way to judge the accuracy of $(\hat{u},\hat{v})$ is to compare the brightness function at a point in the first frame to the brightness function at the predicted new position for the point in the second frame. If disparity is accurately estimated, the brightness values should be similar at these two points. The difference between the two frames

registered by $(\hat{u},\hat{v})$ is an estimate of the directional derivative $I_{\hat{\mathbf{w}}}$ defined in (4.4). We calculate a third estimate of error based upon (4.10) which relates $I_{\hat{\mathbf{w}}}$ to the magnitude of the error in $(\hat{u},\hat{v})$:

$$\varepsilon_{\hat{\mathbf{w}}} = = \left| \frac{I_{\hat{\mathbf{w}}}}{\sqrt{I_x^2 + I_y^2}} \right| \tag{5.17}$$

The error estimate ε_I is converted into a measure of confidence by

$$p_3 = \frac{1}{\varepsilon_I^2 + 1} \tag{5.18}$$

In locations where p_3 is small, the two frames are dissimilar at the predicted place of correspondence and the computed disparity is likely to be in error.

The propagation characteristics of the linear system $\mathbf{G}\hat{\mathbf{w}} = \mathbf{b}$ can be determined by examining the matrix of spatial gradients. Errors on the right hand side of the linear system are magnified by $\|\mathbf{G}^{-1}\|$ in the computed value of disparity. A fourth measure of confidence, based upon the likelihood that errors will be poorly propagated, is given by

$$p_4 = \frac{1}{\|\mathbf{G}^{-1}\| + 1} \tag{5.19}$$

If p_4 is small, then the linear system is ill-conditioned and small measurement errors will tend to produce large errors in $(\hat{u},\hat{v})$.

The four confidence estimates derived above are not independent. The confidences p_1 and p_3 both measure the accumulative error, from all sources, in the disparity estimate. The confidences p_2 and p_4 relate to conditions which are likely to lead to poor estimates: p_2 depends upon a condition which is particularly troublesome for gradient measurement and p_4 conveys the error propagation characteristics of the linear system. Even though the four estimates are not independent we found that they were best treated as separate sources of information and best combined multiplicatively. We examined a number of combination rules and found that the results were

not highly sensitive to the particular rule for combining confidences.

The method for estimating confidence is heuristically motivated. The technique could probably be improved by further examining ways to estimate error and rules for combining several sources of information about errors. The important contribution of this research is to demonstrate the feasibility estimating the accuracy of disparity estimates and usefulness of confidence measurements.

5.1.4.3. Local Optimization with Iterative Registration

The simple method of local optimization can be extended by a method of iterative refinement. In this method disparity estimates obtained from the simple local optimization scheme can be used to register the frame pair. Disparities are then be recomputed with the gradients estimated from the registered frame pair. In section 4.1 it was shown that the measurement error in the temporal gradient could be significantly reduced if the registration reduced the original disparity between the image frames. Since the optical flow field will usually contain variations, the predicted registration will differ across the image. To obtain a consistent linear system, a small region of the first frame must be registered with the second frame on the basis of the predicted disparity at the point for which disparity is to be estimated. A system of linear equations is constructed with gradient constraints line extracted from the registered region.

This process can be performed iteratively, using the disparity estimates at the previous stage to register the frame pair on the next iteration. It is important to emphasize that, at each stage, the registration can only be expected to improve performance if the disparity in the newly registered frame pair is less than the disparity in the previously registered pair. If, for some a point (i,j) if the first frame, the registration is worse than the registration in the last iteration, the new estimate of disparity will, in general, be worse then the previous estimate. It is desirable to register the image only where the disparity estimates are believed to be correct. Therefore, in our

implementation we register in proportion to the confidence in the disparity estimate.

The iterative registration technique is schematically shown in figure 5.6. The registered gradients for the kth iteration at a point (i,j) on the image are estimated as follows,

$$I_x(i,j,t) = \frac{1}{8}\Big[I(i+1,j,t) - I(i-1,j,t) + I(i+1+\hat{u},j+\hat{v},t+1) - I(i-1+\hat{u},j+\hat{v},t+1)\Big] \quad (5.20)$$

$$I_y(i,j,t) = \frac{1}{8}\Big[I(i,j+1,t) - I(i,j-1,t) + I(i+\hat{u},j+1+\hat{v},t+1) - I(i+\hat{u},j-1+\hat{v},t+1)\Big] \quad (5.21)$$

$$I_t(i,j,t) = \frac{1}{2}\Big[I(i+\hat{u},j+\hat{v},t+1) - I(i,j,t)\Big] \quad (5.22)$$

where $(\hat{u},\hat{v})$ is the disparity estimate from the $k-1$st iteration. A flow field of zero disparity vectors is used to initialize the first iteration.

Confidence is estimated as before except that now the changes in the spatial gradients must be calculated with the registered frame pair. The new estimates for the changes in gradients over the registered image pair are,

$$\Delta I_x = \tfrac{1}{2}\Big[I(i+1,j,t) - I(i-1,j,t) - I(i+\hat{u}+1,j+\hat{v},t+1) + I(i+\hat{u}-1,j+\hat{v},t+1)\Big] \quad (5.23)$$

$$\Delta I_y = \tfrac{1}{2}\Big[I(i,j+1,t) - I(i,j-1,t) - I(i+\hat{u},j+\hat{v}+1,t+1) + I(i+\hat{u},j+\hat{v}-1,t+1)\Big] \quad (5.24)$$

The iterative registration technique is employed with variable blurring to produce a coarse-to-fine system of analysis. Images are blurred with a gaussian weighting function. In early iterations the standard deviation of the gaussian weighting function is large. The standard deviation of the weighting function is reduced in each successive iteration. At each level, the radius of the blurring function should be large enough to guarantee that the brightness function is approximately linear over the maximum

expected disparity from the registered images.

The size of the neighborhood from which the constraint equations are selected must depend upon the amount which the images are blurred. At a coarse level of analysis there is little detail which distinguishes nearby points. To obtain sufficiently different constraint equations, the separation between observation points must be increased; otherwise, the conditioning of the linear system will degenerate.

Our system contains four iterations which correspond to four levels of coarseness. The blurring was accomplished by repeated convolution with a 3×3 kernel. The neighbor size and the value of the standard deviation for the approximation to the gaussian weighting function are given in table 5.2 for each of the four iterations.

Iteration	Blur Radius σ	Neighborhood Size
1	7	6
2	5	4
3	3.5	3
4	2	2

Table 5.2

A difficulty with the coarse-to-fine system is that the disparity estimates for stationary and slowly moving points made at coarse levels may be worse than the initially assumed zero vector. To insure that the new disparity estimate made at one level is not worse than the value input into the level, we examine the error bound given by (5.7) for both the initial and new estimates. If the error bound for the new estimate is significantly larger than the bound for the old estimate, it is ignored.

5.1.5. Hybrid Techniques

Matching and gradient-based techniques have different strengths and weaknesses. The performance characteristics of the two techniques are reciprocally related. Matching techniques are capable of producing a sparse sampling of accurately determined disparity vectors. For many applications the vector density produced by matching techniques is insufficient. Gradient-based techniques, on the other hand, generate

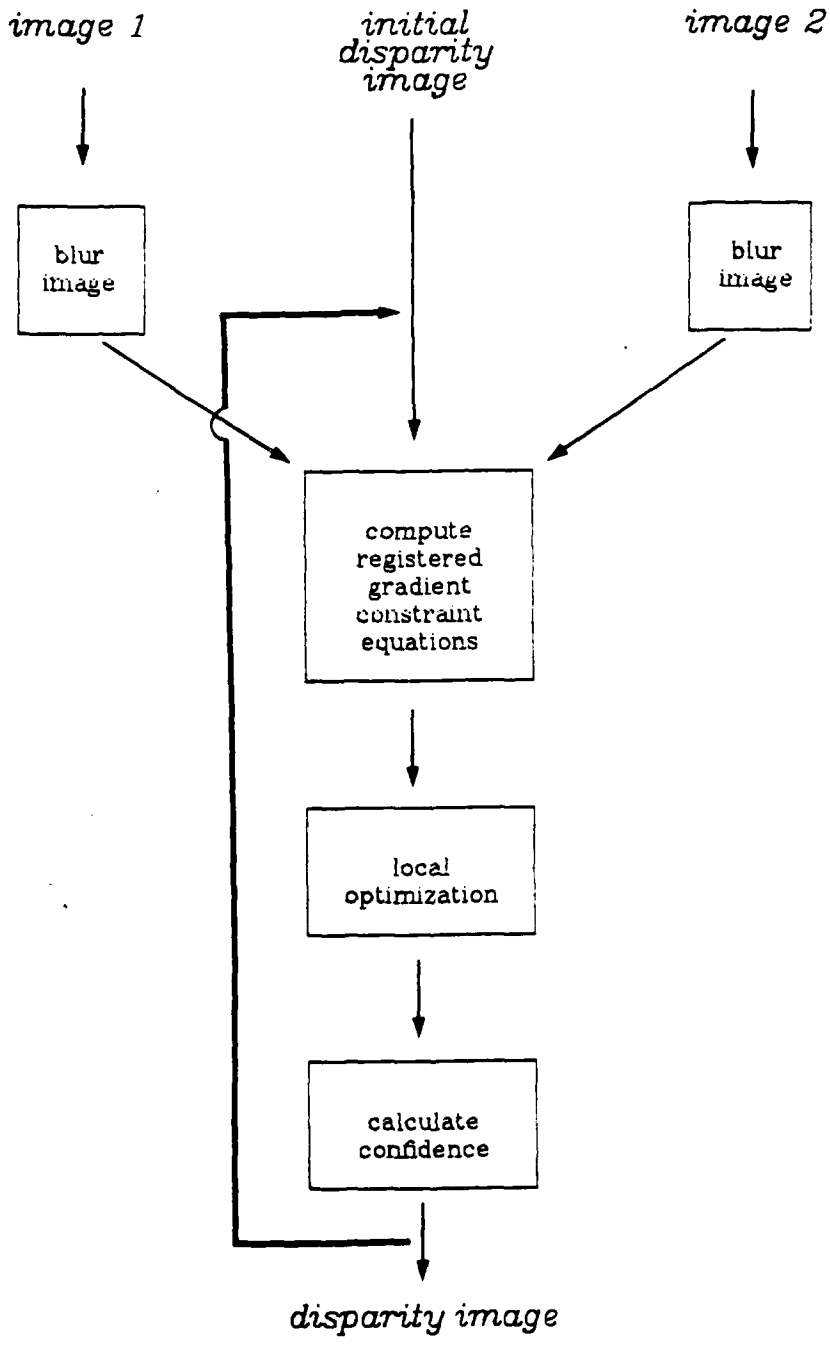


Figure 5.6 Local optimization with iterative registration.

dense fields but are susceptible to a variety of errors. Regions which undergo substantial motion and contain nonlinearities in the brightness function are especially troublesome for gradient-based methods. Gradient techniques tend to work poorly near occlusion boundaries, where the optical flow field is discontinuous. In contrast, matching techniques usually work well at occlusion boundaries as these boundaries are often associated with a distinctive change in surface reflectance. Some improvement can be gained by combining the two methods to take advantage of their different strengths. In this section we approach ways to combine matching and gradient-based techniques to arise at a more robust, hybrid method which takes advantage of the strengths of both matching and gradient-based approaches.

5.1.5.1. Local Averaging

Usually, neighboring points in the image will move in a similar manner. A good prediction of a points motion can be obtained from by examining neighbors whose motion is known. Closer neighbors are, in general, better predictors than are distant neighbors. Thus, disparity at a point can be approximated by the average of nearby estimates, weighted by the distance of the estimates to the point to be approximated. This operation has two effects: the initial estimates are smoothed as they are averaged with other nearby estimates; and the values of previously unknown points are interpolated from the initial estimates. Computationally, the distance-weighted average can be accomplished by a series of local averages.

A serious problem with simple averaging is that disparity values can be inappropriately combined across discontinuities in the optical flow field. Near abrupted changes in disparity, the average of neighbors will usually be a poor predictor of the motions of individual points. Not only will the averaging result in inaccurate interpolation, but the initially correct values will be corrupted by smoothing with points moving in a very different manner. The magnitude of this problem depends, in part, upon the form of

the weighting function. If the variance of the weighting function is very small -- meaning that nearby points are weighted much more heavily than more distant points -- then serious miscalculations can be somewhat limited. However, the density and distribution of the initial estimates must also be considered. Even though nearby points are more heavily waited, if the nearest neighbor lies across a discontinuity in disparity it will provide a poor estimate of motion. Simple distance-weighted averaging, by itself, is not an effective means to generate a dense optical flow field from a sparse sampling of disparity estimates in most situations.

5.1.5.2. Combining Average Motion and the Gradient Constraint

The gradient constraint provides a second source of information about motion. The information available from the local average and the gradient constraint equation are shown graphically in figure 5.7. To combine the gradient constraint with the estimate provided by the motion of nearby point $(\bar{u},\bar{v})$ we place our new estimate on the line perpendicular to the gradient constraint equation which passes through $(\bar{u},\bar{v})$. We expect that the true value of motion will lie between the average of neighbors and gradient constraint equation – on the dashed line segment in figure 5.7. The exact position in which we place the estimate should depend upon the relative confidence which we have in the two sources of information.

Horn and Schunck have developed a method which combines the local average of disparity and the gradient constraint equation [31,37]. Their technique minimizes an error norm based upon departure from smoothness in the flow field -- agreement with the average of neighboring disparity values – and violation of the gradient constraint. The computational method at which they arrive is equivalent to taking a weighted combination of the average of neighboring disparity estimates $(\bar{u},\bar{v})$ and the point (u_p,v_p) on the gradient constraint line where the perpendicular in figure 5.7 intersects the gradient constraint line. The weighting is determined by the magnitude of the spatial gra-

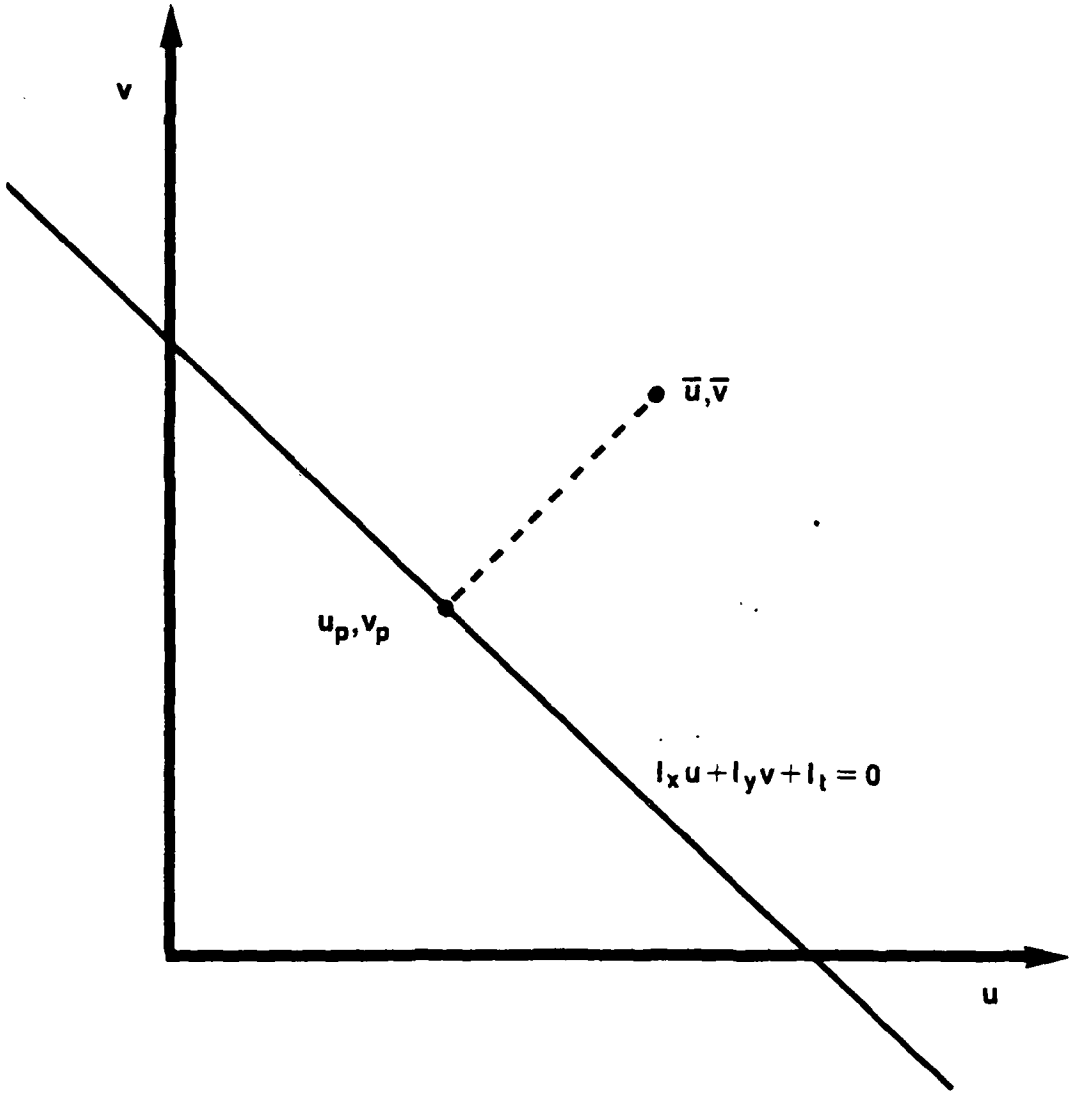


Figure 5.7 The gradient constraint equation and the local average of disparity provide two sources of information amount motion.

dients.

The Horn and Schunck technique can be initialized with a field of zero disparity vectors. If the technique is to be used over a sequence of more than two images, the results of the previous image pair can be used as an initial approximation of the flow field. The method can also be seeded with estimates obtained elsewhere without violating the assumptions of the model [48].

As with simple averaging, the Horn and Schunck technique has difficulty near discontinuities in the optical flow field. Gross miscalculations can be made where the local average is based upon values which lie across a discontinuity in disparity. In section 3.3.1 it was shown that the brightness gradients, on which the gradient constraint equation is based, will be poorly estimated in regions where disparity is rapidly varying. The brightness gradients will also be in error in regions where the brightness function contains nonlinearities and motion is large. The gradient measurement problem is especially serious where surfaces have become occluded, disoccluded, or have left the field of view between frames. Unfortunately, even though the error prone regions are are localized and will usually comprise a small portion of the image, the errors can propagate throughout the image.

5.1.5.3. The Constrained Average

The difficulty encountered with averaging methods is that errors tend to propagate throughout the flow field, even though the problematic regions may be small and localized. If poor estimates can be detected then the affects of the errors can potentially be limited to the regions which are prone to have difficulty. We approach the containment problem by introducing confidence into the estimation process. The local average motion is computed as a the average of neighbors motions weighted by their confidence. Since estimates contribute only in proportion to their confidence "good" estimates which tend to propagate more effectively. A new estimate of disparity

is obtained by combining the average of local disparity estimates and the gradient constraint line as in the Horn and Schunck technique. The constrained average approach is schematically shown in figure 5.8. The frame pair is registered to take advantage of intermediate disparity estimates. The registered gradient constraint equations are computed as in the registered local optimization technique described above.

A confidence is assigned to the new estimate of disparity on the basis of

(1) the likelihood that the local average is in error,

(2) the likelihood that the gradient constraint equation is in error,

(3) the agreement between the average and the constraint equation, and

(4) a bound on the error in the new disparity estimate.

The confidence in the local average is judged by the mean and variance of confidences associated with the estimates which contribute to the average. Estimates contribute to the mean confidence and variance of confidence statistics in the same proportion as they contribute to the average disparity estimate. Two confidences are derived from the weighted average of confidences $\bar{p}$ and the inverse of the weighted variance of confidences σ_p.

$$p_1 = \bar{p} \tag{5.25}$$

and

$$p_2 = \frac{1}{\sigma_p} \tag{5.26}$$

The confidence estimate p_1 and p_2 represent the likelihood that the average of local that the local average of disparity estimates accurately predicts disparity.

The confidence in the correctness of the gradient constraint equation is evaluated by examining the change in the spatial gradients over the sampling interval. Where ΔI_x and ΔI_y are large the gradient constraint equation is likely to be in error. This measure was also used in the local optization method. As before, we estimate the error caused by nonlinearity in the brightness function by the ratio of the change in the spatial gradients to the magnitude of the spatial gradient:

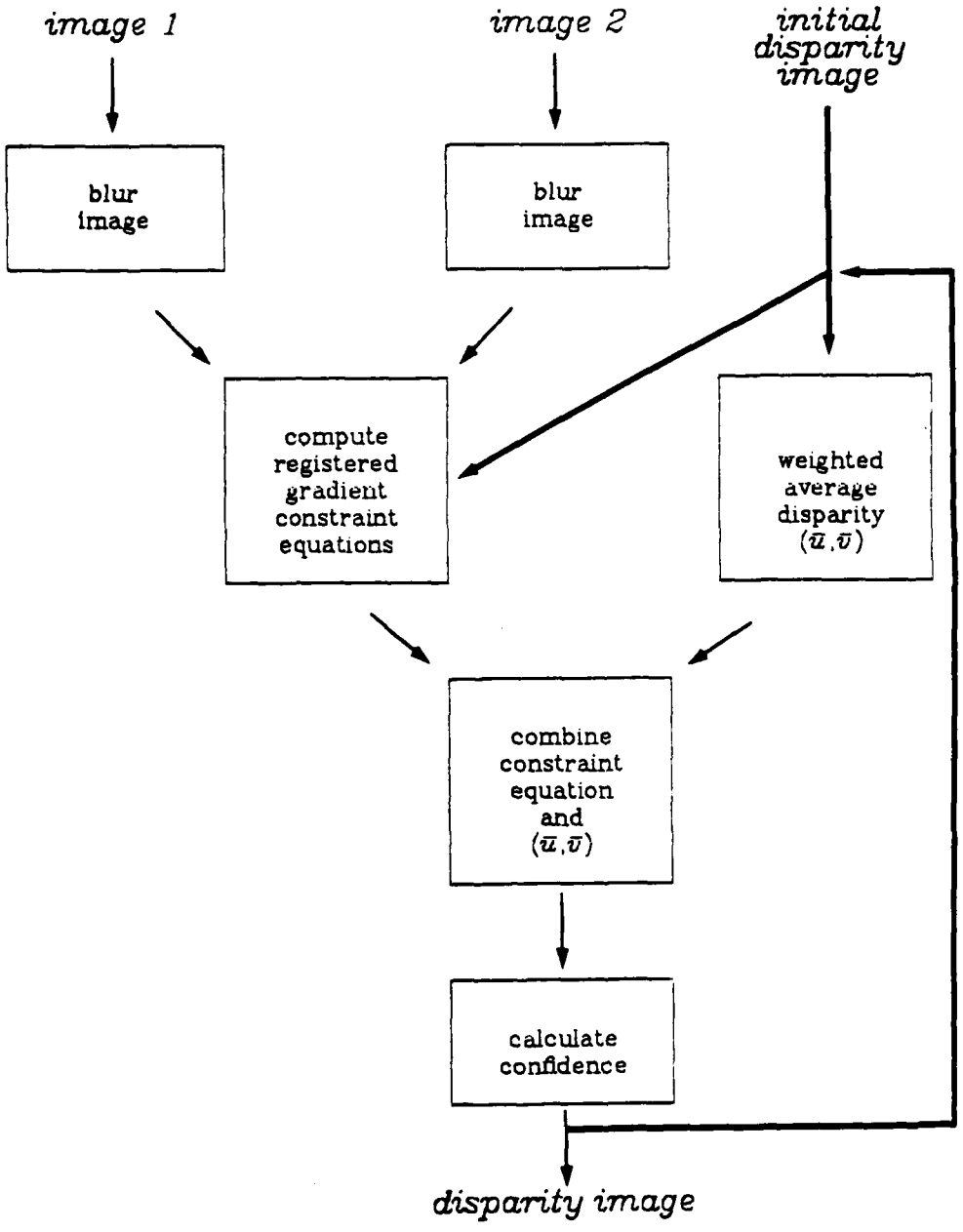


Figure 5.8. A hybrid approach combining an weighted average of neighboring motions and the gradient constraint equation.

$$\hat{\varepsilon}_{\Delta} = \frac{\sqrt{\Delta I_x^2 + \Delta I_y^2}}{\sqrt{I_x^2 + I_y^2}} \tag{5.27}$$

and

$$p_3 = \frac{1}{\hat{\varepsilon}_{\Delta} + 1} \tag{5.28}$$

The confidence value p_3 is sensitive to one condition which causes large errors in the gradient constraint equation.

Inconsistencies between the local average of disparity and the gradient constraint line are detected by examining the relationship between $(\bar{u},\bar{v})$ and the gradient constraint equation. The gradient constraint line should pass through the true value of disparity and, where disparity varies smoothly, $(\bar{u},\bar{v})$ should lie near the true value of disparity. If $(\bar{u},\bar{v})$ is well separated from the gradient constraint line then it can be inferred that one or the other is likely to be in error. A confidence value which indicates the degree of agreement between $(\bar{u},\bar{v})$ and the gradient constraint line is calculated as

$$p_4 = \frac{1}{d} \tag{5.29}$$

where d is the minimum distance between the gradient constraint line and the local average of disparity.

Once a new estimate of disparity has been calculated the difference between the estimated value and the true value can be bounded by,

$$\varepsilon_{\hat{\mathbf{w}}} = = \left| \frac{I_{\hat{\mathbf{w}}}}{\sqrt{I_{x2} + I_y^2}} \right| \tag{5.30}$$

This error bound is used to obtain a confidence in the disparity estimate as

$$p_5 = \frac{1}{\varepsilon_I^2 + 1} \tag{5.31}$$

The confidence p_5 was earlier introduced in the local optimization scheme.

As in the local optimization technique, the separate confidence estimates are multiplicatively combined to arrive at a single estimate of the confidence in the disparity estimate.

5.2. Results

The five methods described above,

(1) coarse-to-fine cross-correlation of feature points,

(2) relaxation feature point matching,

(3) simple local optimization,

(4) local optimization with coarse-to-fine iterative registration, and

(5) gradient constrained averaging initialized with matches obtained with the cross-correlation procedure,

were programmed in the C language on a VAX-11/780. The methods were tested with the two image pairs presented in figure 5.9.a and figure 5.9.b. In the first sequence the camera was stationary. The scene contains a collection of toys. The two trains in the center of the first image move toward each other in the second image. The second sequence simulates a view from an aircraft flying over a city. The images where obtained with a camera fixed on a tripod overlooking a model of downtown Minneapolis. The scene consists of a receding ground plane on which lie a number of structures. The top of the image is furthest from the observer. The camera was moved forward and tilted downward between the first and second frames. Ground truth data is not available for the sequences examined, so our evaluation will only be qualitative.

The results of the coarse-to-fine correlation matching are presented in figures 5.9.c and 5.9.d for the moving trains and simulated flyover, respectively. Disparity vectors are displayed as a white line with a small square box at the vector's base. The correlation program produces a match for every feature point identified by the laplacian feature point selector. This means that incorrect matches will necessarily be made for points which are visible only in the first frame. The correspondences selected for these unmatchable points will, in general, produce low values of cross correlation. Only those points for which the pseudo-normalized cross correlation was quite high ($\rho \geq .99$) are displayed.

(a) moving trains

(b) simulated flyover

(c) correlation matching

(d) correlation matching

(e) relaxation labeling

(f) relaxation labeling

Figure 5.9 Original images and disparity estimates from matching techniques

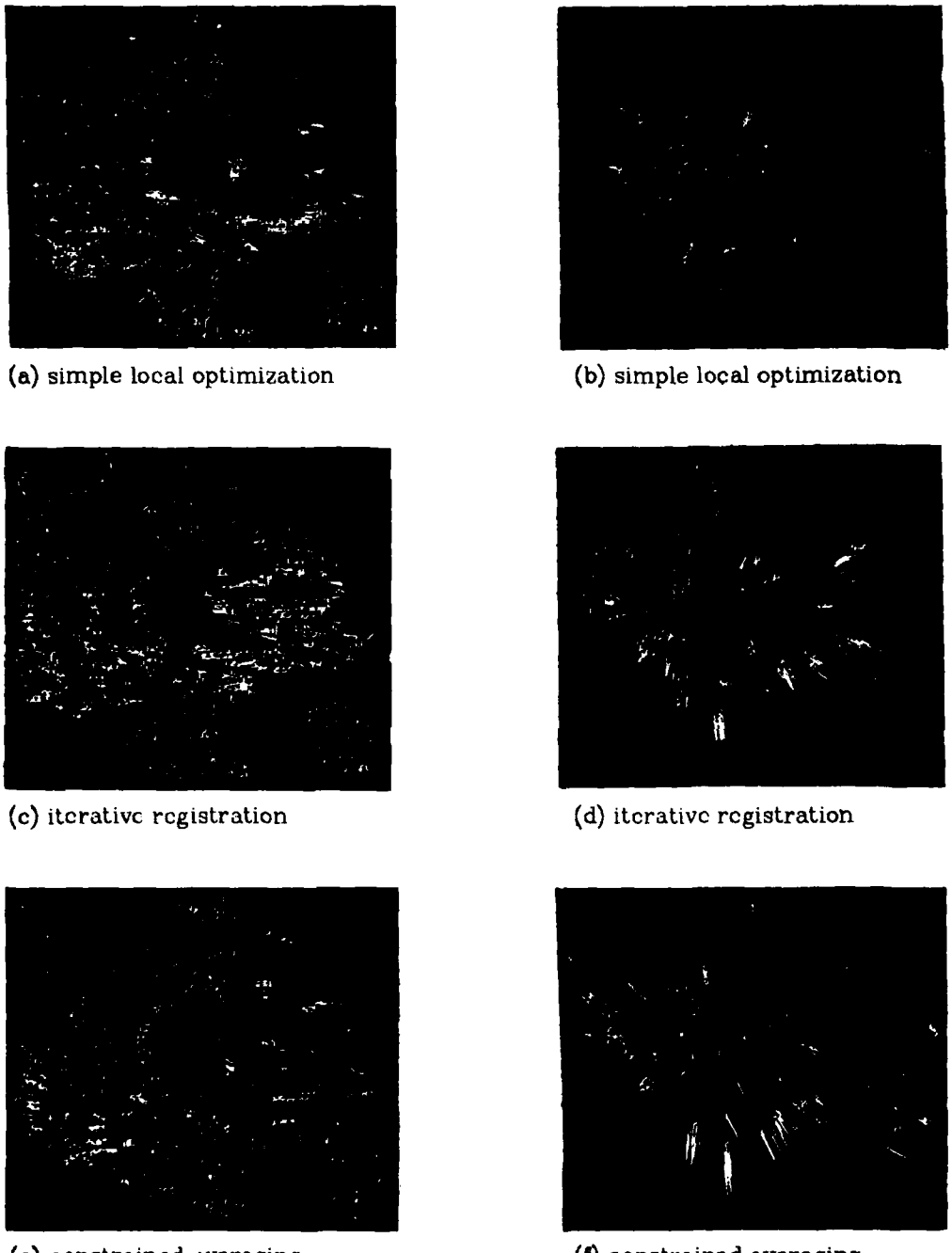
The correlation method produces a sparse sampling of reasonably well distributed points for both sequences. More errors are apparent in the flyover sequence although the percentage of correct disparity vectors in both fields is fairly high. Most of the significant errors tend to be clustered in several small regions.

The disparity fields generated by the relaxation labeling method are shown in figure 5.9.e and figure 5.9.f for the moving trains and flyover sequences. To generate the disparity fields shown here, the label lists for all feature points were examined after ten iterations of the relaxation procedure. Only labels for which the associated confidence was high ($p \geq .6$) were accepted as matches and are displayed in figures 5.9.e and 5.9.f.

The performance of the relaxation technique is comparable to that of the correlation method on the toy trains sequence. However, the two methods produce quite different results for the flyover sequence. The disparity field obtained with the relaxation method is fairly accurate but very sparse. The flyover sequence has some particularly troublesome characteristics for the relaxation approach. The variance of the disparity vectors is high so that neighbors lend little reinforcement. Also, the magnitude of the disparity vectors at the edge of the image approaches the bound on the region from which candidate matches are drawn.

The difference in the performance between the correlation and relaxation techniques on the flyover sequence emphasizes the dependency of these methods upon the problem environment. It is not possible to judge one technique as better than the other on the basis of this difference alone. Before methods can be evaluated, the requirements of the task to be accomplished must be known and specifications developed for the disparity analysis procedure.

Disparity fields obtained with the simple local optimization technique are shown in figure 5.10.a and figure 5.10.b for the moving trains and flyover scenes. Associated with each vector is a confidence in the correctness of the value. A threshold on confidence



(a) simple local optimization

(b) simple local optimization

(c) iterative registration

(d) iterative registration

(e) constrained averaging

(f) constrained averaging

Figure 5.10 Disparity estimates from gradient-based techniques.

was established which produced a reasonably dense sampling of mostly correct values. Only vectors which exceeded the confidence threshold are displayed. The resulting field was too dense to clearly display the entire field. Consequently, only 25% of the vector field is shown in figures 5.10.a and 5.10.b.

The results of the coarse-to-fine method of iterative refinement are shown in figures 5.10.c and 5.10.d. The disparity fields obtained with the hybrid method which combines the cross correlation approach with gradient constrained smoothing are displayed in figures 5.10.e and 5.10.f. The hybrid estimates were the result of 16 iterations of the constrained averaging technique described above. On each iteration the local average was computed over a 5x5 window, centered on the point to be estimated, for every point in the image. Confidence thresholds were established for both the method of iterative registration and the hybrid technique which produced vector densities for the moving trains scene which were comparable to that obtained with simple local optimization. The numeric values of the confidence thresholds are not meaningful by themselves. The confidence values are obtained in a different manner in the three different gradient techniques which are demonstrated here. The confidences are not normalized and hence the values can not be compared across methods. The disparity fields in figures 5.10.c, 5.10.d, 5.10.e, and 5.10.f are also subsampled versions of the actual fields -- only 25% of the vectors are shown.

The disparity fields obtained with the gradient-based techniques are all substantially more dense than the fields produced by the matching techniques. All of the gradient-based techniques produce reasonably accurate results with the moving train sequence. The simple local optimization method seems to generate more errors with the moving train scene than either of the other gradient-based techniques. The method of iterative registration produces a field which is generally correct with a small number of errors interspersed through most of the field. The best results are obtained with the hybrid technique; the few errors which are evident occur in 3 or 4 small

regions.

The three techniques are more easily distinguished on the basis of their performance with the flyover sequence. The simple local optimization method produces a large number of errors even for the relatively sparse sampling of vectors displayed in figure 5.10.a. The method of iterative registration and the hybrid technique generate many fewer errors in fields which are much more dense than that obtained with the simple local optimization approach. Again, the most accurate results appear to be obtained with the hybrid approach, although it is difficult to judge the magnitude of the difference between the hybrid and iterative registration techniques without ground truth data.

Note, for the gradient-based techniques, the areas where very few vectors are displayed. Disparity is poorly estimated in these regions and low values of confidence are assigned to the estimates obtained there. The problematic regions are usually fit into one or more of the following characterizations:

1. largely homogeneous regions,
2. highly textured regions which are moving, or
3. regions which contain large discontinuities in the flow field.

Disparity estimates obtained in homogeneous areas are likely to be in error because of the poor propagation characteristics of linear systems constructed in these regions. The temporal gradient is poorly measured in highly textured regions which undergo significant motion. In regions which contain large discontinuities in the flow field the temporal gradient is poorly estimated and the systems of equations from the region are likely to contain inconsistencies.

It is interesting to compare the disparity fields produced by the hybrid technique and the cross correlation method. Recall that the output of the cross correlation technique was used to initialize the constrained averaging. The gradient constrained smoothing serves two purposes: (1) to fill out the sparse set of estimates obtained with

cross correlations and (2) to increase the overall accuracy of estimates in the vector field. An increase in density is clearly evident when comparing figures 5.9.c and 5.10.e (remember that only 25% of the vector field is displayed in figure 5.10.e). The hybrid method also appears to be significantly less prone to error. The improvement in accuracy accrues, in part, from the averaging of estimates and the additional constraint of the gradient relationships. These factors act to produce better estimates. Another reason for the improvement is the power of the heuristics, developed earlier, to judge the quality of the estimates. Not only are better estimates obtained but the confidence value associated with the estimates allows the opportunity to select the appropriate trade-off between accuracy and density.

The success with which confidence estimates predict the accuracy of disparity estimates is demonstrated in figures 5.11.a, 5.11.b, 5.11.c, and 5.11.d. The disparity field produced by the simple local optimization technique with the moving trains sequence is displayed in with a low threshold on confidence in figure 5.11.a and a high threshold in figure 5.11.b. As before, only 25% of the vectors which exceed the threshold are displayed. Similar thresholds are shown for the method of iterative registration in figures 5.11.c and 5.11.d. Only 15% of the vectors which exceed the thresholds set for figures 5.11.c and 5.11.d are displayed. For both methods confidence provides a reasonable index of the accuracy of disparity estimates. A sparse sampling of accurate estimates exceeds the high confidence threshold. When the threshold is lowered, more dense fields are obtained with a significantly greater number bad vectors.

Confidence is based upon a number of heuristics which identify conditions which are likely lead to errors or indications that an estimate is in error. The usefulness of the residual error in detecting errors made by the simple local optimization technique is demonstrated in figure 5.11.e. This field was obtained with the moving trains sequence. Only vectors which poorly satisfy the system of constraint equations on which the estimate was based are displayed. Most of the vectors displayed are poor

(a) simple local optimization
(low threshold)

(b) simple local optimization
(high threshold)

(c) iterative registration
(low threshold)

(d) iterative registration
(high threshold)

(e) simple local optimization
(vectors with high residual)

Figure 5.11 **Detecting errors with confidence.**

estimates of the actual disparity.

5.3. Summary

The results presented here highlight the intrinsic differences between matching and gradient-based methods. Both matching approaches produced sparse samplings of generally accurate vectors. The disparity fields obtained with the gradient-based techniques are much more dense than those produced by matching methods. The increase in density is achieved with little or no decrease in accuracy. Furthermore, by providing reliable confidence estimates, gradient-based methods provide an added flexibility which is not available with matching techniques; the choice to trade density for accuracy is made explicit.

The dependency of performance upon the nature of the scene is evident from the characteristics of the disparity fields obtained with the two different scenes. Before any particular technique is selected for a task, the methods must be studied within the problem environment.

The gradient-based techniques which were implemented demonstrate the feasibility of measuring the quality of disparity estimates. More study is required to better understand the heuristics by which confidence is estimated. Gradient-based techniques are susceptible to a variety of problems and tend to produce very poor estimates in troublesome areas of the image, as is shown in figure 5.11.e. Without accurate estimates of confidence, good estimates can not be distinguished from bad and gradient-based techniques are of little use. With accurate confidence estimates, poor disparity estimates can be filtered from the field and information can be propagated from areas of high information content into areas of low information content.

The results show the improvement to be gained over simple local optimization by iterative registration and coarse-to-fine analysis, particularly when motions are large as in the flyover sequence. However, the best overall performance was obtained with

the hybrid technique which combines correlation matching and gradient constrained smoothing. The hybrid approach produced reasonably dense samplings of disparity vectors for both sequences with a high level of accuracy.

6. CONCLUSIONS

A number of conclusions can be draw from this study. It is important to emphasize, however, that this was a feasibility study and thus conclusions are only tentative.

First of all, it is clear to us that disparity estimation techniques can not be developed without careful consideration of the interpretive processes that will depend on the estimates. Clear trade-offs exist between different aspects of performance. The optimal mix of performance criteria will be application dependent. This application dependence must be more precisely described before design specifications for a disparity estimation system are developed. If generality across applications is desired, then we strongly suspect that a variety of different algorithms will be required. (In fact, there is evidence that the human visual system uses two quite different and independent process to determine motion induced disparity [49].)

There are intrinsic limits to the precision with which disparity can be determined. In a sense, all of the disparity estimation techniques that we have investigated depend on the same sort of information in image sequences. These sequences frequently contain large areas where disparity cannot be unambiguously determined. This is particularly true for large homogeneous regions or regions which contain only parallel linear patterns. Image features such as noise also limit the accuracy of disparity estimation. It is important to understand the nature and causes of these limits in order both to get the best estimates possible and to design interpretation processes in a sensible manner.

The computational requirements of real-time disparity estimation are substantial. Processing rates in excess of 100 million operations per second may well be required. If very high accuracy of individual estimates is required, at least some of these operations will likely involve floating point computations. The state of the art in high performance computer architectures strongly suggests that data rates of this sort require

highly structured, pipelinable algorithms. This is reinforced by the serial nature of the outputs of most image sensors. Both the gradient and correlation techniques fit into this form. Token matching systems involving any sort of cooperative computation do not. We do not expect that such algorithms will be implementable in real-time unless and until highly parallel architectures with complex interconnection networks are available. Even the prototyping of such architectures is still some time away.

For those tasks requiring reasonably high density and dispersion of disparity estimates, the gradient-based algorithms appear to be a promising choice. We have performed an analysis of the limits on accuracy for such methods and have shown how improved performance can be obtained in a number of situations. We have also shown that the gradient techniques work best when an initial approximation is available for disparity estimates. This suggests that a hybrid techniques be investigated in which a sparse, initial estimate is obtained using cross-correlation and then a final, denser estimate is obtained from a gradient-based method.

7. COMPUTER ARCHITECTURES FOR DISPARITY ESTIMATION

7.1. Overview

A detailed architectural analysis of possible implementations of disparity estimation techniques must wait until a more complete evaluation of potential algorithms has been completed. It is possible, however, to make some relatively general observations that may assist in choosing among algorithms of comparable performance.

All of our algorithm simulations have been performed on a medium capacity general purpose computer (a VAX-11/780). Simulation performance has in general been several orders of magnitude slower than "real-time". Processing times vary, depending on the algorithm, from a fraction of a minute to many minutes for a single frame pair. For a given form of disparity analysis, processing speeds can only be improved through faster processing elements or the utilization of parallel computations. The VAX is capable of performing on the order of one million computational operations a second. The use of very high speed general purpose processing elements would yield on the order of a ten fold performance increase, though at substantial expense. Additional speed ups must come from some form of parallelism. Even with substantial parallelism, however, obtaining real time performance will be difficult. For example, a hundred fold increase in processing throughput will typically require significantly more than one hundred processors operating in parallel with algorithms carefully tailored to support the parallelism.

7.2. Quantifying Performance

The concept of *real-time* performance must be precisely quantified in the design specifications for any disparity analysis system implemented in hardware. Informally, real-time performance implies that over an extended period of time, the device can operate at a speed compatible with the incoming image data. That is, the *throughput*

of the device must match the incoming data rate. A more precise specification requires a complete task description. The incoming data rate must be specified with respect to number of pixels per frame, number of frames per second, and average and peak disparities within each frame. The quality of the input signal is also important as increased noise will lead to greater processing requirements.

Throughput is not the only important characterization of processing speed. Not only must the input data rate be accommodated, but any significant change in the input must be signaled in the output within a reasonable amount of time. Thus, *latency* effects must also be included in the specifications. Again, this becomes task specific. Some tasks require very short response times. For others, somewhat longer response times can be tolerated. Throughput and latency requirements often work against one another. Throughput can be increased by using pipelining, but long pipelines will increase the latency of the system.

Throughput may also be increased by exploiting the temporal redundancy in image sequences. In most environments, motion is relatively constant. Knowing the image dynamics at one point in time provides a reliable estimate of future changes in the image. As a result, many dynamic properties need not be computed for each frame pair, but instead can be computed over a much longer sequence of frames. For example, several of the gradient based techniques require many iterations to produce reliable results. These iterations can be performed repeatedly over the same frame pair. However, if motions are relatively constant, each succeeding iteration can be done over the next frame pair in the sequence. In fact, this produces an additional benefit of averaging out the effects of uncorrelated image noise. Dependence on this effect has important performance implications. First, the task specification must be such that it is reasonable to assume relatively constant motion. Should motions change significantly at some point, many frame pairs will be required for the system to reconverge. Thus, latency will increase significantly when motions change.

Finally, the nature of sensors used by the system may have architectural implications. In particular, almost all current sensors provide data in a raster format. A linearized byte stream is produced corresponding to some particular scanning pattern on the image. As a result, architectures well suited to raster processing may prove more efficient that those using some other organization such as SIMD parallelism.

7.3. Parallelism

Parallel processing in image understanding is possible because many computational processes can be performed in a independent manner at many image locations simultaneously. Thus, these operations (or portions of the operations) can be executed on separate processors. The type of parallel architecture is determined by the way in which the computational operations are partitioned out onto separate processors and the data rate, topology, and synchronizations required for interprocessor communication.

7.3.1. Pipeline Architectures

Pipelining is possible when spatially distinct and independent computational operations consist of a linear sequence of sub-operations. With pipelining, the parallelism occurs over this sequence of sub-operations, not over spatial position in the image. Separate processors are assigned to each sub-operation. Data is passed from one processor to the next in sequence. When one processor in the pipeline is finished with a particular computation at a particular image location, it can start on the same computation at the next location without having to wait for the completion of processing elements farther down the pipeline. Ten processing elements in a pipeline can thus result in a ten-fold increase in throughput, provided all elements perform tasks of comparable complexity.

For a DIDA system, two forms of pipelining are possible. On a small scale, many of the possible algorithms depend on structured computations consisting of vector operations such as dot products. Such computations are particularly well suited for pipelining. In fact, many commercial *vector processors* use pipeline architectures specifically intended for dot product style computations. Furthermore, this small scale pipelining is well suited to several of the signal processing architectures being developed for the VHSIC program. On a larger scale, many disparity estimation algorithms consist a sequence of complex but independent computational procedures. For example, the gradient techniques involve blurring each image, finding gradients, solving constraint equations, filtering out inconsistent results, and, possibly, iterating the whole process with more refined initial estimates of disparity. Each of these procedures could be implemented as a step in a pipelined computational system.

7.4. Processor Arrays

Image understanding computations can also be partitioned by dividing up the problem into spatially distinct components and then executing each component on a separate processor. Because this collection of processors now has a direct geometric correspondence to the original image, this organization is often referred to as a *processor array*. The primary limitation of processor array architectures comes from problems with inter-processor communications and synchronization. Typically, spatially distinct operations are not in fact entirely independent. Processor arrays are severely limited in the complexity of interconnections that can be implemented. If the algorithm requires a substantial amount of interaction between processors, any potential improvement in processing power can be lost to communications overhead. This is particularly true for smaller scale parallelism in which a relatively few number of processors is each responsible for computations over a relatively large portion of each image. The problem may be less severe, at least for some tasks, as the number of pro-

cessors increases and, correspondingly, the function of each processor is simplified. In the limiting case, it is likely to be desirable to have one processor for each pixel in the image. Unfortunately, this is well beyond the state of the art in terms of our ability to fabricate processors. At the present time we have only a limited understanding of how to design image processing algorithms for such an architecture. In addition, this large scale parallelism would require parallel access to input data, another feature not currently realizable.

7.5. Architectural Considerations for Gradient Techniques

The gradient based techniques are well suited to parallel implementation. As they involve substantial amounts of computation, the exploitation of this potential parallelism is crucial to real-time implementation. The following paragraphs give an overview of some of the architectural possibilities for computing gradient-based disparity estimation devices. A more complete analysis should wait until more precise performance criteria are developed and appropriate techniques are selected.

The local optimization techniques consist of three major steps:

1. Compute the gradient constraint equations.
2. Calculate the solution to the linear system of local constraint equations.
3. Estimate the confidence for the disparity estimate.

All these operations can be independently performed at each position in the image. As a result, they can easily be decomposed into a pipelined organization. Accuracy can be improved by a variety of iterative refinement techniques in which disparity estimates obtained at one step of the process are used to re-register local image regions and then the estimation process is repeated. It may also be desirable to modify the blurring function during the iterative process and so implement a coarse-to-fine analysis. To perform iterative refinement of estimates the output of each stage of processing must be fed back into the system. If the processing elements of the system are dupli-

cated, this can be done in a pipelined manner. A sequence of identical processors would be connected such that each processor performed one step in the iteration.

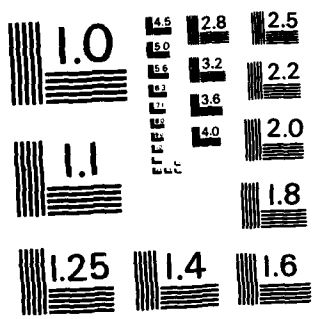
Gradient-based smoothing consists of four major steps:

1. Compute the average of neighboring disparity estimates.
2. Calculate a confidence in the average disparity estimate.
3. Compute the gradient constraint equation.
4. Combine the average estimate and the gradient constraint equation to arrive at a new estimate for disparity

This sequence is repetitively performed to smooth a sparse sampling of points obtained from the cross correlator. Here again, a sequence of operations is performed and the operations are limited to small independent neighborhoods of the image. Thus, pipelining is straightforwardly implemented.

Within each of the operational steps discussed above, there are highly structured numeric computations. Gradient estimations, solutions to over-determined linear systems of equations, residual vector computations, and local averaging can all be described in terms of a limited set of vector and matrix operations. These operations can be implemented in high speed pipelined hardware by exploiting the highly structured nature of matrix arithmetic. Thus, major portions of the techniques can be decomposed into pipelinable components, each of which can be implemented with pipelined arthimetic operations.

AD-A128 327 DIDA - DYNAMIC IMAGE DISPARITY ANALYSIS(U) MINNESOTA 2/2
UNIV MINNEAPOLIS DEPT OF COMPUTER SCIENCE
W B THOMPSON ET AL. 31 DEC 82 AFWAL-TR-83-1035
UNCLASSIFIED F33615-81-K-1541 F/G 9/2 NL



8. RECOMMENDATIONS

We recommend that the next stage of the DIDA project concentrate primarily on the development of design specifications sufficiently precise so that continued efforts at system development will be focused in productive directions. Development of these specifications requires continued activity in three related areas: Specification of goal dependencies, criteria selection, and data base generation. Work in these areas will lead to meaningful design specification and also aid in the study of new and improved algorithms initiated during the current contract period.

8.1. Criteria Selection

The successful prototyping of hardware for real-time disparity analysis depends on the existence of precise and realistic preformance criteria. These criteria must be specified with a view towards utility, feasibility, and verifiability.

The criteria under which a particular algorithm can be evaluated have been informally described above. These criteria need to be formally specified in any continued effort. *Accuracy* specifies the average precision of the disparity field. Accuracy should be separately specified with respect to magnitude and orientation. *Density* specifies the number of points in the image assigned a disparity value. *Scene dependency* specifies how accuracy varies as a function of different scene properties (eg. accuracy near object edges vs. accuracy in the center of large surfaces). *Start-up and hysteresis* specify the behavior of the algorithm on the first few frames of a sequence or when major changes occur in scene dynamics. *Graceful degradation* specifies the effectiveness of an algorithm as boundary conditions are approached with respect to noise, maximum disparity, or difficult scene types.

8.2. Goal Dependency

Certain performance criteria that are important for particular scene types and interpretation tasks are relatively unimportant for other types of scenes and interpretation. It is extremely important that follow-up work describe a range of realistic *scenarios* for which disparity analysis might prove useful. A set of problem domains should be described informally with respect to sensor type, the nature of objects and surfaces, and possible object and sensor dynamics (and thus properties such as resolution and maximum disparity). These descriptions should be produced with the advice and assistance of the technical monitor at the Avionics Laboratory. The necessary requirements for at least four interpretive tasks should be specified: map matching, 3-D model matching, obstacle avoidance, and segmentation.

8.3. Data Base

In order to evaluate disparity estimation techniques, a standard data base of relevant image sequences is required, along with the "correct" interpretations of those sequences. The data base should contain long image sequences covering all of the problem domains and interpretation tasks that are a part of relevant scenarios. Real imagery will require that "truth" data be acquired through other sensors or some form of interactive analysis. Synthetic imagery must be generated in a manner which avoids artifacts that effect the performance of analysis algorithms. It is suggested that AFWAL/AAA take the lead in this activity.

8.4. Demonstration of Algorithms

In order to gain experience with both the evaluation process and disparity estimation algorithms with the potential for real-time implementation, we suggest that four state-of-the-art algorithms be demonstrated with respect to the formal evaluation criteria. Based on our experience during the current contract period, we recommend

that algorithms to be demonstrated include temporal-spatial gradient analysis, token matching, image matching (cross-correlation), and a technique combining gradient analysis and token matching. Once the utility of the evaluation process has been demonstrated, additional algorithms, possibly written by a variety of research groups, should be tested. Again, this activity should be centered at AFWAL/AAA.

8.5. Algorithm Development

Work under the current contract has successfully produced an analytical examination of the intrinsic limitations of several state-of-the-art algorithms. This analysis is being used to modify these algorithms so as to improve performance. We are particularly interested in a number of hybrid approaches which can combine the strengths of several different existing approaches. We hope to continue both analytical and empirical studies for algorithm development.

BIBLIOGRAPHY

[1] C. Cafforio and F. Rocca, "Methods for Measuring Small Displacements of Television Images," *IEEE Trans. Information Theory*, vol. IT-22, pp. 573-579, September 1976.

[2] J.O. Limb and J.A. Murphy, "Estimating the velocity of moving images in television signals," *Computer Graphics and Image Processing*, vol. 4, pp. 311-327, December 1975.

[3] C.L. Fennema and W.B. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Processing*, vol. 9, pp. 301-315, April 1979.

[4] D.B. Gennery, "Object detection and measurement using stereo vision", *Proc. 6th Int. Joint Conf. on Artificial Intelligence*, pp. 320-327, August, 1979.

[5] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Computer Graphics and Image Processing*, vol. 7, pp. 195-210, April 1978.

[6] J.L. Potter, "Velocity as a cue to segmentation," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-5, pp. 390-394, May 1975.

[7] J.L. Potter, "Scene segmentation using motion information," *Computer Graphics and Image Processing*, vol. 6, pp. 558-581, December 1977.

[8] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283-287, Oct. 15, 1976.

[9] S. Ullman, *The Interpretation of Visual Motion*, Cambridge: MIT Press, 1979.

[10] S.T. Barnard and W.B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-2, pp. 333-340, July 1980.

[11] T.D. Williams, "Depth for camera motion in a real world scene," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-2, pp. 511-516, November 1980.

[12] C.J. Jacobus, R.T. Chien, and J.M. Selander, "Motion detection and analysis by matching graphs of intermediate-level primitives," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-2, pp. 495-510, November 1980.

[13] M.J. Hannah, "Computer matching of areas in stereo images," AIM-239, Stanford University, July 1974.

[14] J.A. Leese, C.S. Novak, and V.R. Taylor, "The detection of cloud pattern motions from geosynchronous satellite image data," *Pattern Recognition*,

vol. 2, pp. 279-292, December 1970.

[15] E.A. Smith and D.R. Phillips, "Automated cloud tracking using precisely aligned digital ATS pictures," *IEEE Trans. Computers*, vol. C-21, pp. 715-729, July 1972.

[16] D.I. Barnea and H.F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Trans. Computers*, vol. C-21, p. 179-186, February 1972.

[17] H.-H. Nagel, "Formation of an object concept by analysis of systematic time variations in the optically perceptible environment, *Computer Graphics and Image Processing*, vol. 7, pp. 149-194, April 1978.

[18] R. Jain and H.-H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-1, pp. 206-214, April 1979.

[19] R. Jain, W.N. Martin, and J.K. Aggarwal, "Segmentation through the detection of changes due to motion," *Computer Graphics and Image Processing*, vol. 11, pp. 13-34, September 1979.

[20] R. Nevatia, "Depth measurement by motion stereo," *Computer Graphics and Image Processing*, vol. 5, pp. 203-214, June 1976.

[21] W.B. Lacina and W.Q. Nicholoson, "Concept validation of depth aided target acuisition for the cruise missile," Northrop Research and Technology Center, November 1978.

[22] W.F. Clocksin, "Perception of surface slant and edge labels from optical flow: a computational approach," *Perception*, v. 9, pp. 253-269, 1980.

[23] O. Firschein, M.J. Hannah, D.L. Milgram, and C.M. Bjorklund, "*Passive Imagery Navigation*," Technical Report LMSC-D076313, Palo Alo Research Laboratory, Lockheed Missiles and Space Company, Inc., Palo Alto, CA., December, 1980.

[24] H.P. Moravec, "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," Ph.D. dissertation, Stanford University, 1980.

[25] D.B. Gennery, "A stereo system for an autonomous vehicle," *Proc. 5th Joint Int. Conf. Artificial Intelligence*, Cambridge, MA, Aug. 1977, pp. 576-582.

[26] S. Ullman "Analysis of Visual Motion by Biological and Computer Systems," *Computer* v. 14, no. 8, pp. 57-69, August 1981.

[27] K.M. Mutch and W.B. Thompson, "Hierarchical Estimation of Spatial Properties from Motion," Technical Report 82-22, Dept. of Computer Science, University of Minnesota, 1982.

[28] R.L. Lillestrand, "Techniques for change detection," *IEEE Trans. Computers*,

vol. C-21, pp. 654-659, July 1972.

[29] "Special Issue on Computer Architecture for Patern Analysis and Image Database Management," *IEEE Transactions of Computers*, v. C-31, no. 10, October, 1982.

[30] P. R. Wolf, *Elements of Photogrammetry (with air photo interpretation and remote sensing)*, McGraw-Hill Book Company, New York, 1974.

[31] B.K.P. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[32] R.J. Schalkoff, "Algorithms for a real-time automatic video tracking system," Ph.D. Dissertation, University of Virginia, May 1979.

[33] A. N. Netravali and J. D. Robbins, "Motion-Compensated Television Coding: Part I," *The Bell System Technical Journal*, Vol. 58, No. 3, March, 1979.

[34] W.B. Thompson and S.T. Barnard, "Low-level estimation and interpretation of visual motion," *Computer*, August 1981.

[35] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pp. 674-679, August, 1981.

[36] R. J. Schalkoff and E. S. McVey, "A Model and Tracking Algorithm for a Class of Video Targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI 1-4, No. 1, pp. 2-10, January, 1982.

[37] B. G. Schunck and B.K.P. Horn, "Constraints on optical flow," *Proceedings IEEE conference on Pattern Recognition and Image Processing*, pp. 205-210, Aug. 1981.

[38] G.E. Forsythe and C.B. Moler, *Computer Soluion of Linear Algebraic Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967.

[39] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applicaitons to Image Analysis and Automated Cartography," *Communications of the ACM* Vol. 24, pp. 381-395, June 1981.

[40] H.P. Moravec, "Towards automatic visual obstacle avoidance," *Proc. 5th Int. Joint Conf. on Artificial Intelligence*, p. 584, August 1977.

[41] H.P. Moravec, "Visual mapping by a robot rover", *Proc. 6th Int. Joint Conf. on Artificial Intelligence*, pp. 598-600, August 1979.

[42] J.L. Crowley, "A representation for visual information," The Robotics Institute, Carnegie-Mellon University Technical Report CMU-RI-TR-82-7, 1982.

[43] L. Dreschler and H.-H. Nagel, Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," *Proceedings of the 5th International Joint Conference on Artificial*

Intelligence, pp. 692-697, August, 1981.

[44] D. Marr and T. Poggio, "A theory of human stereo vision," *Proc. Royal Soc. London*, vol. 204, pp. 301-328, 1979.

[45] D. Marr and E. Hildreth, "Theory of edge detection," MIT AI memo 518, April 1979.

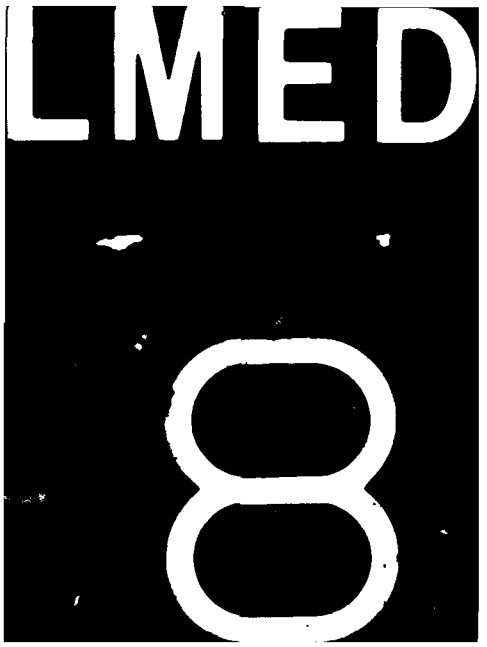
[46] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond.*, v. B 207, pp. 187-217, 1980.

[47] S.T. Barnard, "The image correspondence problem," Ph.D. dissertation, Computer Science Department, University of Minnesota, December 1979.

[48] M. Yachida, "Determining Velocity Maps by Spatio-Temporal Neighborhoods from Image Sequences," submitted for publication, 1982.

[49] D.A. Marr, *Vision*, San Francisco: W.H. Freeman and Company, 1982.

*USGPO: 1983-659-062-797


LMED
8